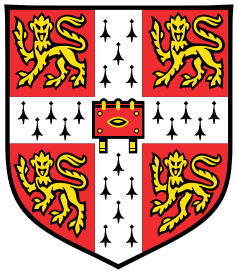
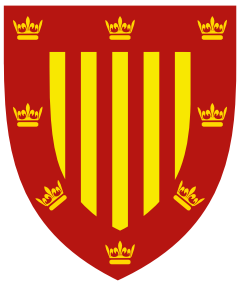
Cosmology from galaxy clusters with cosmic microwave background lensing mass calibration





Íñigo Zubeldía Lafuente

Institute of Astronomy and Peterhouse
University of Cambridge

This thesis is submitted for the degree of
Doctor of Philosophy

April 2020

Dedicated to my parents

Declaration

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared below and specified in the text. It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared below and specified in the text. I further state that no substantial part of my thesis has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared below and specified in the text. It does not exceed the prescribed word limit for the relevant Degree Committee. Those parts of this thesis which have been published or submitted for publication are as follows:

- Chapters 2 and 3 have been published as:
  Íñigo Zubeldia, Anthony Challinor, Cosmological constraints from Planck galaxy clusters with CMB lensing mass bias calibration, *Monthly Notices of the Royal Astronomical Society*, Volume 489, Issue 1, October 2019, Pages 401–419.

- Chapter 4 has been accepted for publication in MNRAS as:
  Íñigo Zubeldia, Anthony Challinor, Quantifying the statistics of CMB-lensing-derived galaxy cluster mass measurements with simulations.

Íñigo Zubeldía Lafuente
Tudela, April 2020

Summary

Cosmology from galaxy clusters with cosmic microwave background lensing mass calibration

In this thesis, we present a cosmological analysis of the galaxy clusters in the *Planck* MMF3 cosmology sample, which consists of 439 Sunyaev-Zel'dovich-detected clusters, with a cosmic microwave background (CMB) lensing calibration of the cluster masses. As demonstrated by *Planck*, galaxy clusters detected through their SZ signature offer a powerful way to constrain cosmological parameters such as $\Omega_{\rm m}$, which parametrises the mean matter density of the Universe, and σ_8, which characterises the amplitude of the matter perturbations. Determining the absolute cluster mass scale is, however, difficult, and some recent calibrations have yielded cosmological constraints in apparent tension with constraints in the ΛCDM model derived from the power spectra of the primary CMB anisotropies.

In order to calibrate the absolute mass scale of the full *Planck* cluster sample, we measure the CMB lensing signals of 433 of its clusters (those with measured redshift) with *Planck* temperature data. We calibrate the bias and intrinsic scatter of our CMB lensing mass observable, the CMB lensing signal-to-noise, with mock observations from an N-body simulation. We then perform a joint likelihood analysis of the cluster counts and mass observables taking as input the CMB lensing signal-to-noise ratios, SZ signal-to-noise ratios, and redshifts. Our analysis uses a likelihood that properly accounts for selection effects in the construction of the cluster sample. We find $\sigma_8 \left(\Omega_{\rm m}/0.33\right)^{0.25} = 0.765 \pm 0.035$, $\Omega_{\rm m} = 0.33 \pm 0.02$, $\sigma_8 = 0.76 \pm 0.04$, and $1 - b_{\rm SZ} = 0.71 \pm 0.10$, where the mass bias factor $1 - b_{\rm SZ}$ relates cluster mass to the SZ mass that appears in the X-ray-calibrated cluster scaling relations. We find no evidence for tension with the *Planck* primary CMB constraints on ΛCDM model parameters.

Acknowledgements

First, I would like to express my most sincere gratitude to my supervisor, Anthony Challinor. I would like to thank him for guiding me through my PhD with his always insightful comments and suggestions and with his careful examination of my ideas and my writings. I am deeply grateful, too, for the support and encouragement that he has always given me throughout my time as his student.

I would like to extend my gratitude to my first, second, and third year assessors, Martin Haehnelt, Anthony Lasenby, Debora Sijacki, and Colin DeGraf, for their helpful comments and advice. I would also like to thank James G. Bartlett, Jean-Baptise Melin, Ewald Puchwein, William Handley, Antony Lewis, Sebastian Bocquet, George Efstathiou, Steven Gratton, William Coulton, and Oliver Friedrich for useful discussions about my research, and Anna Bonaldi and Richard Battye for providing me with the *Planck* SZ noise estimates, essential for my work. My gratitude also goes to Max Pettini for his excellent Part II cosmology course, which I had the pleasure to supervise, and to my college tutor, Christopher Lester, for his help with college-related matters.

Pursuing a PhD would have not been possible without the financial support of my funding bodies, the Isaac Newton Fund and the Isbel Fletcher Studentship Fund, to which I am very thankful. I am grateful, too, to my college, Peterhouse, which financially supported my attendance to two academic conferences, and to the IoA for its student conference fund and for granting me three months of post-submission funding.

Having Antón and George as office mates has been a true privilege (my conversations in Spanish with Antón will be particularly missed), as has been sharing the celebrated IoA coffee breaks with Lewis (to whom I am also indebted for his invaluable help with my many computer-related issues), Lukas, Aneesh, Roger, Tom, and many others. Special mention goes to the many post-lunch coffees with Lukas and Lewis at Churchill. The cosmological and non-cosmological conversations we had there will remain one of my personal highlights of my time in Cambridge.

Outside of the IoA, I am fortunate to have met some great people in my college, with whom I have attended more dinners and post-drinks than I would like to admit. I would like to thank Peerapat, in particular, for our enjoyable trips to museums and to the opera. I should also mention my friends from my years in Madrid, Miguel, Carlos Enol (our meetings in London helped me escape occasionally the Cambridge bubble), and Iván (for our pints and formals at Wolfson and Peterhouse), and those from my hometown, Víctor (who has helped me more than once with his computer skills), Sergio and Chema. Our enduring friendship has been a phenomenal source of support over the past three and half years.

Finally, I would like to acknowledge the unfailing support and love of my parents, Jesús and Begoña, and my sister, Laura. This thesis would have not been possible without them.

Table of contents

List of figures

List of tables

Chapter 1

Introduction

1.1 Galaxy clusters as cosmological probes

1.1.1 The observed Universe

The last century has witnessed spectacular progress in our knowledge of the Universe. Decades of careful observations have revealed a fascinating cosmos full of complexity yet, under the right optics, remarkably simple. There has also been a huge effort to understand what we see from a theoretical point of view, a quest that has sometimes yielded paradigm-shifting theories, and that remains ongoing. It was just over a hundred years ago that Einstein developed his ground-breaking theory of gravity, general relativity (GR; Einstein 1916), which superseded the then well-established theory of Newton (Newton, 1687), and which remains, unchallenged by observations, our best theoretical description of gravity (see, e.g., Will 2014 and Ishak 2018 for recent reviews of tests of GR). At that time, it was not clear whether the faint, diffuse *nebulae* seen in the sky, systematically studied since the times of Messier and Herschel in the 18th century and known today as galaxies, were part or our Galaxy or were separate 'island universes', the latter a point of view that dates back to at least Kant (Kant, 1755). This question was definitively settled with Hubble's discovery of cepheids in the Andromeda Galaxy in the 1920s. Cepheids are variable stars whose pulsation period is related to their luminosities, and which can therefore be used as 'standard candles' to measure distances to their host galaxies. Using them, Hubble showed that Andromeda was too far to be within our Galaxy (Hubble, 1929b). Shortly after, Hubble also famously discovered the recession of distant galaxies, thus providing the first solid piece of evidence for the expansion of the Universe (Hubble, 1929a). This possibility had already been explored theoretically by Friedmann, Lemaître, and others

(Friedmann, 1922, 1924; Lemaître, 1927), who showed that a dynamic Universe is a natural consequence of Einstein's GR.

These discoveries marked the beginning of cosmology as a science. The understanding of nuclear physics in the ensuing years resulted in the development of the hot Big Bang theory, according to which the Universe started as a very dense, hot plasma from which most of the light elements in the Universe (mostly hydrogen and helium) formed (Alpher et al., 1948, 1953). This theory was put on a firm observational footing with the detection in the 1960s of the relic radiation that decoupled form this early plasma when its temperature became low enough and that has been free-streaming ever since, the cosmic microwave background (CMB) (Penzias and Wilson, 1965). The last decades of the 20th century and the first ones of the 21st century have seen cosmology coming of age and becoming a precision science. Observations of distant Type Ia supernovae, which can be used as standard candles, first gave evidence for the acceleration of the expansion of the Universe in the late 1990s (Riess et al., 1998; Perlmutter et al., 1999). Cosmologists have now measured the spectrum and the tiny anisotropies in the CMB with exquisite accuracy, constituting the current gold standard of observational cosmology (Mather et al., 1994; Bennett et al., 1996; Hinshaw et al., 2013; Planck 2018 results I, 2018). In parallel, the distribution of matter in the Universe has also been studied with a number of different approaches with great success (e.g., Tegmark et al. 2004; Eisenstein et al. 2005; Cole et al. 2005; Fu et al. 2008; Abbott et al. 2019). There has been great progress in the theoretical front too, with the general features of how large scale structures form being well understood (Lifshitz, 1946; Peebles and Yu, 1970; Zel'Dovich, 1970; Press and Schechter, 1974; Gott and Rees, 1975; White and Rees, 1978; Bardeen, 1980; Peebles, 1980; Kodama and Sasaki, 1984; Liddle and Lyth, 2000). An important contribution to this effort has come from the development of numerical simulations, which are now highly sophisticated and can reproduce rather well how the Universe looks on large scales (Klypin and Shandarin, 1983; Davis et al., 1985; Evrard, 1988; Navarro et al., 1997; Cole et al., 2000; Springel et al., 2005; Tinker et al., 2008; Klypin et al., 2011; Vogelsberger et al., 2014).

The picture that emerges from this enormous effort throughout the last century is the following. We live in a Universe that is remarkably homogeneous and isotropic on very large scales, roughly over 200 Mpc, and that has a spatial curvature consistent with zero. It is expanding, it has always been expanding (we believe), which implies that its age is finite (about 13.8 billion years), and the expansion is currently accelerating. At present time, matter from the Standard Model of particle physics (i.e., that formed by protons, electrons, etc., known in cosmology as 'baryonic matter') only constitutes

about 5 % of the energy content of the Universe. Dark matter, a form of matter (or, more broadly, an entity that, on astrophysical and cosmological scales, behaves like pressure-free matter) whose interaction with the rest of the components of the Universe is compatible with being only via gravity, further accounts for about 25 % of this energy content. This dark matter is 'cold', i.e., has negligible velocity dispersion. The remaining 70 % is accounted for by dark energy, a mysterious component that is responsible for the acceleration of the expansion and that, as of today, is compatible with Einstein's famous cosmological constant, Λ, which can be seen as some form of energy density associated with the vacuum.

The homogeneity and isotropy of the Universe on very large scales contrasts with the wealth of structure that is seen on smaller scales across many orders of magnitude, from the realms of quantum physics to planets, stars, and galaxies. On the largest scales, matter is distributed forming the cosmic web, comprising filaments and sheets, nodes at the intersections of filaments (galaxy clusters), and huge underdense regions in between (voids). A simulation of the cosmic web can be seen in Figure 1.1, in which a slice of the 250 Mpc box of the Bolshoi simulation is shown; the filaments, nodes, and voids comprising the cosmic web can be easily identified. The cosmic web originated mostly through the gravitational evolution of initially very small density fluctuations, the early Universe being remarkably smooth. These primordial fluctuations, which are almost scale-invariant, are believed to have been originated as quantum fluctuations during an initial phase of rapid accelerated expansion known as cosmic inflation.

This general picture has become the 'standard' or 'concordance' model of cosmology, also known as the ΛCDM paradigm (Λ standing for the cosmological constant, and CDM, for cold dark matter), which remarkably can be summarised with only a few parameters. With it, cosmologists can describe all current observations[1]. This model, however, is an 'effective' description far from being completely satisfactory. Indeed, dark matter, despite the broad evidence for its existence, from galactic scales to the CMB, has so far eluded direct detection in a laboratory, and little is is known with certainty about its physical nature (see, e.g., Schumann 2019 for a recent review). Dark energy is even more enigmatic: thus far, attempts to calculate theoretically its energy density as some kind of energy density associated with the vacuum have famously failed by many orders of magnitude (e.g., Carroll 2001), and trying to envision it as something more than just a cosmological constant has only yielded null results (see, e.g., Huterer and Shafer 2018 for a review). Finally, cosmic inflation, despite its success on several fronts (see Section 1.2.2), cannot be said to be completely proven by evidence,

[1]Except, perhaps, the ongoing 'H_0 tension'; see, e.g., Knox and Millea (2019) for a review.

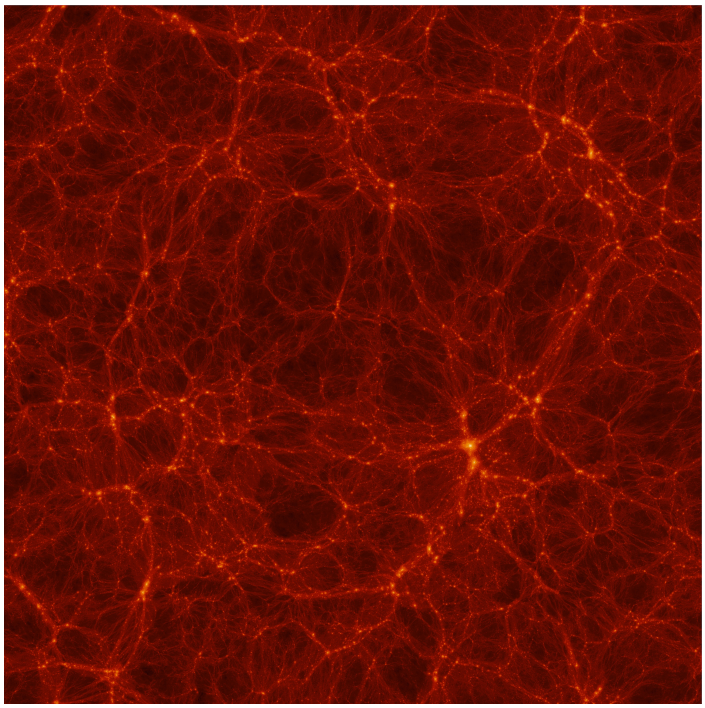


Fig. 1.1 Slice from the 250 Mpc box of the Bolshoi simulation at $z = 0$, showing the cosmic web, with its nodes (galaxy clusters), filaments, and voids. Figure credit: Klypin et al. (2011).

and its microphysics is not well understood. Indeed, one needs to invoke new physics to make it happen, with currently dozens of different inflation models proposed in the literature, none of which has been detected in a laboratory. This and the next decades, with an array of new experiments currently under development or planning, will hopefully shed light on these lesser-understood aspects of ΛCDM, perhaps even revealing the need for a different paradigm.

Galaxy clusters, the main focus of this thesis, have historically played an important role in establishing this standard picture of the Universe, and will continue to be of foremost cosmological interest in the future (Allen et al., 2011; Dodelson et al., 2016). There are two main ways in which galaxy clusters are useful for cosmology. First, their abundance as a function of mass and redshift is a sensitive probe of cosmology. Following this idea, in Chapter 3 we constrain several cosmological and astrophysical parameters with galaxy clusters detected by the *Planck* satellite. Second, the amount of baryonic matter in galaxy clusters relative to the total mass is expected to reflect the mean baryonic abundance in the Universe, and can be used to determine it, among other applications.

In the next sections of this chapter, we will review what galaxy clusters are, how they form, how we can observe them and, most importantly, the details about how we can infer cosmological information from them.

1.1.2 Galaxy clusters in context

What are galaxy clusters?

It is a truth universally acknowledged, that galaxy clusters are the largest gravitationally bound objects in the Universe[2]. They are part of the cosmic web, the largest scale structure in the Universe, sitting at the intersections of filaments. They have masses ranging from about $10^{13} M_\odot$ (galaxy groups) to about $10^{15} M_\odot$ (the most massive clusters in the Universe), and a typical average density of a few hundred times the mean density of the Universe (Allen et al., 2011). Thus, they occupy a relatively small fraction of the cosmic volume, being the peaks of the matter distribution and of the gravitational potential on large scales. They form due to gravitational collapse of the rarest peaks of the initial density field, which was possibly set up by inflation. This process is complex and extended in time: clusters grow and evolve with time as a result of a sequence of mergers with other clusters, and of accretion of smaller systems, preferentially along filaments (Kravtsov and Borgani, 2012).

The matter content of galaxy clusters can be divided into three different components. First, they contain galaxies, with a typical cluster hosting about 100–1000 of them. These are mostly elliptical and lenticular galaxies with little ongoing star formation, especially in central regions; a large, bright elliptical galaxy often sits at the cluster centre (Kravtsov and Borgani, 2012). In addition, there is hot, diffuse, ionised gas that fills the cluster potential well in a smooth manner, not being associated to the individual galaxies. This gas constitutes the *intra-cluster medium* (ICM). It consists mostly of ionised hydrogen and helium, reflecting the cosmic abundance, with traces (about 1 % in mass) of heavier elements, some of them only partially ionised (Werner et al., 2008). Despite being very hot, the ICM is very low density, with typical number densities of around $10^{-1}\,\mathrm{cm}^{-3}$ in the cluster centre, and $10^{-5}\,\mathrm{cm}^{-3}$ in the outskirts (Allen et al., 2011). Finally, there is dark matter, which constitutes the *dark matter halo*.

Galaxy cluster is something of a misnomer, for most of the cluster mass, about 80–90 %, is in the dark matter halo (Allen et al., 2011). This reflects the mean cosmic abundance, which is perhaps not surprising given the cosmological scales of clusters.

[2]For the sake of concision, we do not provide any references to support this statement.

Dark matter is therefore responsible for most of the cluster's gravity; as we will see, this makes it relatively easier to understand clusters (their origin, their internal structure, their abundance, and their distribution) from a theoretical point of view. This also means that the most important quantity governing a cluster's properties is its mass. Galaxies only account for a small fraction of the baryonic matter: the stellar mass is about 20 % of the mass in the ICM for group-sized clusters with masses of about $10^{13} M_{\odot}$, and about 5 % for massive clusters with masses of about $10^{15} M_{\odot}$ (Kravtsov and Borgani, 2012). Galaxy formation in clusters is, thus, a very inefficient process.

Overview of observational signatures

As multi-component beasts, galaxy clusters can be observed across different parts of the electromagnetic spectrum. At the high-frequency end are X-rays. The ICM is very hot, with typical temperatures of about 10^7–10^8 K. This is consistent with the virial temperature expected for a cluster typical mass, i.e., the temperature of the gas if its particles are in virial equilibrium within the potential well of the cluster, $T_{\rm vir} \sim GM\mu/k_{\rm B}R$, where M is the cluster mass, R is its radius, and μ is the ICM gas mean molecular weight (Kravtsov and Borgani, 2012). This thermal energy has its origin in the gravitational energy of the ICM gas, which is converted into thermal energy as the gas falls into the cluster potential well and thermalises. As a consequence of these high temperatures, the ICM emits strongly in X-rays, mostly through thermal *bremsstrahlung*, or free-free radiation, due its highly ionised state, with smaller contributions from free-bound emission, and from bound-bound line emission from the partially-ionised ICM metals (Allen et al., 2011). This emission makes galaxy clusters easily recognisable in X-ray observations: they are the only bright, continuous, spatially extended (typically a few arcmin), extragalactic X-ray sources. The other relevant bright extragalactic X-ray sources are point-like active galactic nuclei (AGNs), which can be easily distinguished from resolved clusters (Allen et al., 2011).

X-ray observations allow the measurement of clusters' X-ray luminosity (most commonly in the soft band, about 0.5–2 keV), L_X, and of their temperature (see, e.g., Böhringer and Werner 2010 for a review). These quantities can be measured as a function of angular separation from the cluster centre with precise observations, typically by averaging over concentric annuli, or integrated within some aperture (or within a spherical volume, for which a model is needed), as is often the case in surveys in which many clusters are considered (Kravtsov and Borgani, 2012). The X-ray luminosity is a direct probe of the cluster gas number density, n_g, since $L_X \propto n_g^2$. Thus, the integrated X-ray luminosity is a measure of the total cluster gas mass and,

under certain assumptions, of the total cluster mass. In addition, detailed observations of the luminosity and the temperature as a function of angular separation from the centre enable, under the assumption that the cluster is in hydrostatic equilibrium, the determination of the cluster enclosed mass as a function of radius, as we will see in Section 1.3.

In the optical and near-infrared, most of the cluster emission is starlight from its galaxies. Thus, at these frequencies a cluster looks like a collection of independent galaxies (indeed, as a *cluster of galaxies*), and can be identified by looking for overdensities in the distribution of galaxies. Quantities of interest that are typically considered in optical surveys, and which correlate with the cluster total mass, are the richness (the number of galaxies within a given angular aperture), the total luminosity, and the colour (Allen et al., 2011). More detailed studies of individual clusters can also measure the galaxy number density and the galaxy velocity dispersion as a function of angular separation from the centre, and the velocity dispersion of the individual galaxies. As we will see in Section 1.3, assuming dynamical equilibrium it is possible to obtain the cluster total mass as a function of radius with these two quantities.

In addition to the optical and near-infrared light emitted by the cluster galaxies, light emitted by external galaxies that happen to be behind the cluster can also be studied, and this is in fact a very interesting thing to do. The reason is gravitational lensing (see, e.g., Bartelmann 2010 for a review). A consequence of GR is that not only does gravity affect massive bodies, but it also affects light. The trajectory of light emitted by, e.g., a galaxy that lies behind a massive object (e.g., a galaxy cluster) is indeed deflected due to the gravity of such object, which acts as a *lens*. This results in the source object having distorted observed size and shape. In the case of galaxy clusters, the main regime of interest is that of weak lensing, in which this effect is small and in which the key quantity controlling it is the matter distribution of the lens integrated along the line of sight, the lensing convergence. In this regime, a cluster's gravity slightly modifies the observed ellipticity of background galaxies in a correlated way, which makes it possible to determine the cluster lensing convergence and, if a model is assumed, its mass (see Section 1.3). Lensing is completely insensitive to the cluster's dynamical state, and thus so are lensing mass estimates. This property has turned galaxy lensing into the current 'gold standard' of galaxy cluster mass estimation. Indeed, mass estimates from other techniques (e.g., from X-rays) do depend on the cluster's dynamical state, which, as we will see, can cause significant biases in the estimated masses, which, in turn, propagate into biases in cosmological constraints obtained with them (e.g., Pratt et al. 2019).

At CMB, or mm, wavelengths, the Sunyaev-Zel'dovich (SZ) effect due to galaxy clusters can be observed (Sunyaev and Zeldovich 1972; see, e.g., Rephaeli 1995a, Carlstrom et al. 2002, and Mroczkowski et al. 2019 for reviews). When the CMB photons travel through a galaxy cluster, some of them interact with the free electrons of the ICM via inverse Compton scattering. Since the ICM electrons are very hot and the CMB is very cold, there is a net positive transfer of energy from the ICM to the CMB photons. Photons are, on average, upscattered towards higher frequencies. This produces a spectral distortion over the original CMB blackbody spectrum, with an intensity decrement at frequencies below $\sim 217\,\mathrm{GHz}$, and an increment above this crossover frequency. This effect is known as the thermal SZ (tSZ) effect; it is illustrated in Figure 1.2, in which the spectral distortion due to an extremely massive cluster is shown. The tSZ effect can be observed with multi-frequency observations in the mm; in fact, its characteristic spectral signature makes it relatively easy for clusters to be found in mm surveys. Provided that they can be resolved, clusters appear as small, extended (a few arcmin), smooth objects clearly distinct from the larger, approximately Gaussian CMB anisotropies. In addition, they appear as colder than the CMB at lower frequencies (below $217\,\mathrm{GHz}$), and hotter than it at higher frequencies. The amplitude of this spectral distortion, the Compton-y parameter, y_{SZ}, is proportional to the cluster gas pressure integrated along the line of sight, which makes the tSZ effect an interesting probe of the ICM. Moreover, the integrated Compton-y parameter within some aperture (or some volume, assuming a model) correlates tightly with the cluster total mass. Since this amplitude is the measure of a spectral distortion, it does not decrease with distance, unlike, e.g., X-ray or optical luminosities. This makes the tSZ effect a very interesting cluster observable, especially valuable to study clusters at high redshift.

Another interesting observable at CMB frequencies is CMB lensing. Similarly to light from background galaxies, the CMB photons are also deflected by galaxy clusters (and, indeed, by the whole LSS of the Universe; see, e.g., Lewis and Challinor 2006 for a review). This provides another way to determine cluster masses with CMB observations. As in galaxy lensing, these estimates are free from any assumptions about the cluster dynamical state, which makes them very valuable. It also has the advantage that the CMB is behind all the clusters in the observable universe, which makes CMB lensing particularly interesting for high-redshift clusters. Nevertheless, unlike galaxy lensing, using CMB lensing to determine cluster masses is in its infancy, with current estimated mass signal-to-noise ratios per cluster smaller than unity. However, the mean mass of a given cluster sample can be determined to high significance, which means

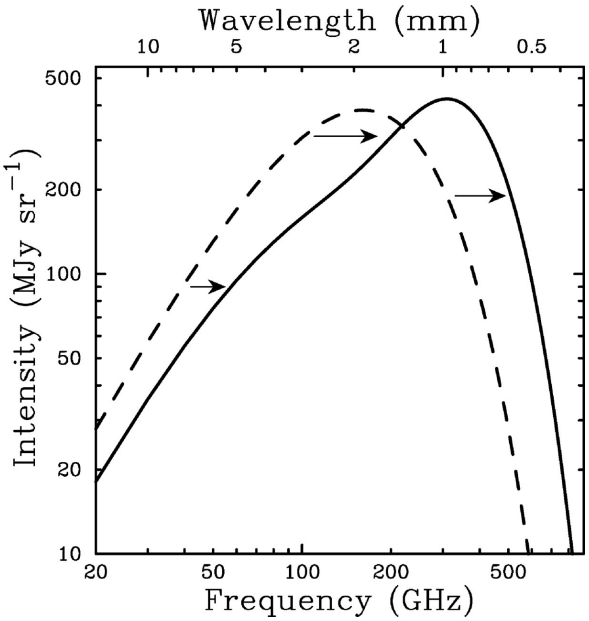


Fig. 1.2 Illustration of the spectral distortion due to the thermal SZ effect. The dashed curve is the undistorted CMB blackbody spectrum, whereas the solid line is the tSZ-distorted spectrum that would be caused by a cluster 1 000 times more massive than the typical galaxy cluster. Figure credit: Carlstrom et al. (2002).

that CMB cluster lensing can already be useful for cosmology (see, e.g., Chapters 2 and 3). Galaxy cluster CMB lensing is reviewed in some detail in Section 1.4, and its use in order to estimate galaxy cluster masses is one of the main contributions of this thesis (see Chapters 2, 3, and 4).

Figure 1.3 shows the multi-component, multi-wavelength appearance of a simulated massive cluster taken from the Illustris simulation (Vogelsberger et al., 2014). The upper left panel shows the dark matter distribution, with the main cluster halo in the centre, and with smaller halos associated to the cluster's galaxies around it. External halos can also be seen, mostly distributed along filaments. The upper right panel shows the gas distribution, which is smoother than the dark matter distribution, except for the small, dense regions within each galaxy. The middle left panel depicts the cluster metallicity, with higher metallicity areas corresponding to galaxies. The middle right panel shows the cluster as it would look in the optical, in which its galaxies are seen. Note the correspondence of galaxies with dark matter halos and with peaks in the gas density, and also note the large elliptical galaxy sitting at the cluster centre. Finally, the two bottom panels show the cluster's X-ray appearance and its tSZ signal (left and right panels, respectively); it can be seen that both signals reflect the gas distribution, with the X-ray signal more centrally concentrated than the tSZ signal.

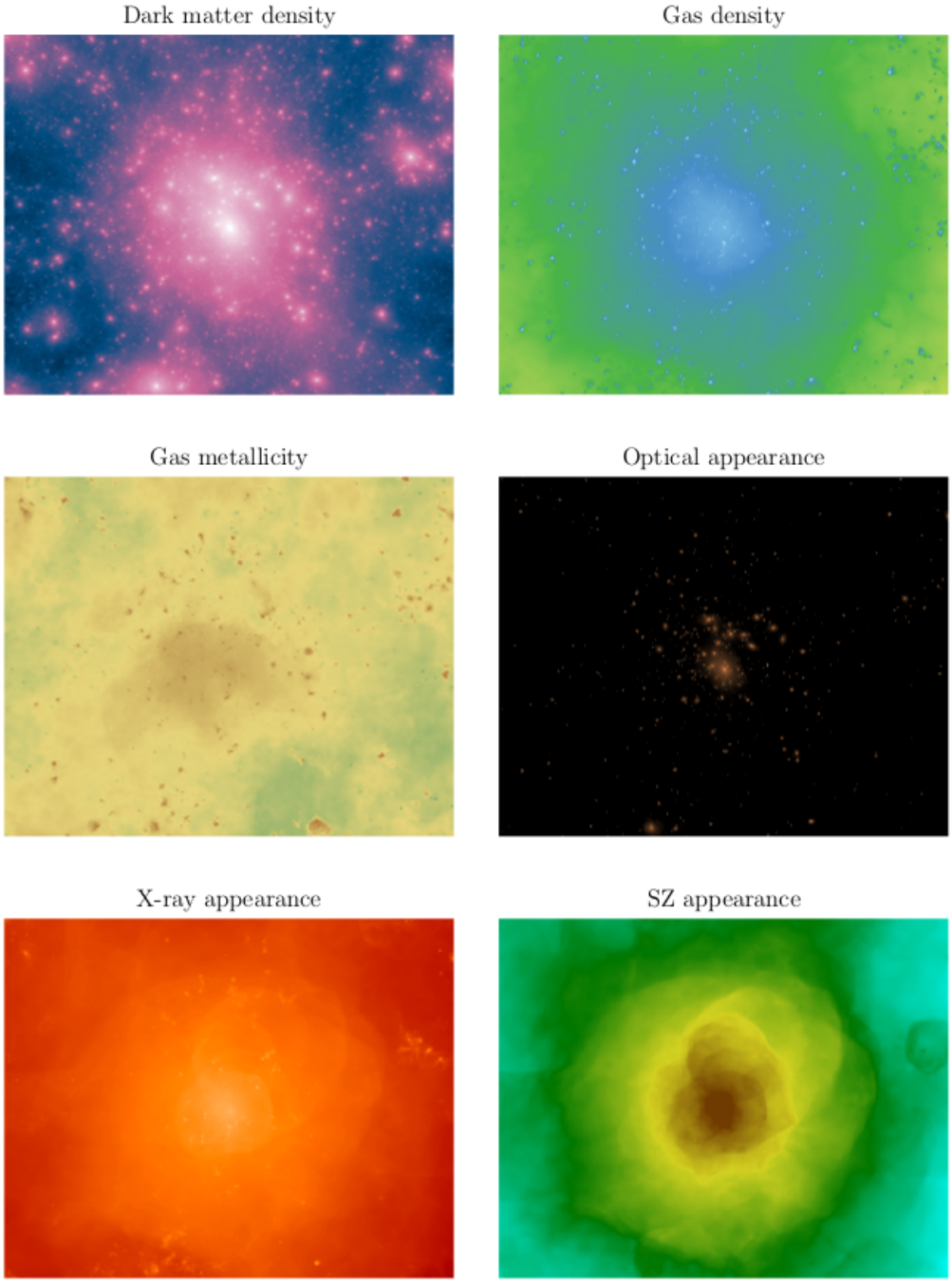


Fig. 1.3 Massive galaxy cluster from the Illustris simulation (Vogelsberger et al., 2014). The different panels show the cluster's dark matter density, gas density and gas metallicity distributions, and how the cluster would look in optical, X-ray, and mm (SZ) observations. Figure credit: `www.illustris-project.org`.

Fig. 1.4 X-ray and optical emission from the Coma cluster, a massive galaxy cluster at $z = 0.0231$ containing over 1 000 galaxies. Figure credit: `sci.esa.int`; ESA/XMM-*Newton* (X-ray), SDSS (optical).

For comparison, a real cluster, the massive Coma cluster, is shown in Figure 1.4. The diffuse X-ray emission, as seen by XMM-*Newton*'s European Photon Imaging Camera (EPIC), is shown along with the optical light from the cluster's galaxies as photographed by the Sloan Digital Sky Survey (SDSS).

Galaxy cluster surveys: a historical perspective

There is a long history of observing galaxy clusters and of using them in order to deepen our knowledge of the Universe (see Biviano 2000 for an enjoyable review of the history of galaxy cluster science). Indeed, although it was not until the 1920s that the extragalactic nature of galaxy clusters was confirmed, observations of the clustering of *nebulae* date back to, at least, the late 18th century. It was in 1784 that Messier, famous for his catalogue of *nebulae*, which contains 103 of them, 30 of which are now known to be galaxies, reported a concentration of *nebulae* in the Virgo constellation (Messier, 1784), perhaps the earliest reference to a galaxy cluster (the Virgo cluster). A year later, in 1785, Herschel described what we now know is the Coma cluster in the following terms: '*that remarkable collection of many hundreds of nebulae which are to be seen in what I have called the nebulous stratum of Coma Berenices*' (Herschel, 1785). He catalogued about 2 500 *nebulae* and identified several other clusters in the

process, a task that would be continued by his son in the 19th century, who identified over 6 000 *nebulae*. Several other catalogues of *nebulae* were elaborated throughout the 19th century, a landmark being Dreyer's *New General Catalogue of* nebulae (Dreyer, 1888), which by 1908 contained over 13 000 of them.

The realisation, in the 1920s, that galaxies are external to the Milky Way, which was then definitively recognised as just another galaxy, revealed the enormous size of clusters. This arguably constituted the starting point of modern galaxy cluster science. Shortly after, Zwicky famously applied the virial theorem to the Coma Cluster, finding that its galaxies were moving too fast for the system to be stable, which lead him to propose that a large part of its mass was in the form of invisible, dark matter (Zwicky, 1937). (The existence of dark matter, however, would not be widely accepted until the 1970s.) A couple of decades later, the first modern catalogues of clusters were produced, with the landmark Abell catalogue being published in 1958 (Abell, 1958).

All the observations mentioned above, and the catalogues compiled with them, were, of course, optical. The development of X-ray astronomy in the 1970s revealed the hot, X-ray emitting ICM (Cavaliere et al., 1971; Gursky et al., 1971; Meekins et al., 1971), which had already been proposed some years earlier (Limber, 1959; van Albada, 1960). Since then, many X-ray catalogues have been produced and, since the early 1990s, used in cosmological studies. A milestone was the launching of the ROSAT satellite in 1990, which performed an all-sky survey of the X-ray sky, RASS (Rosat All-Sky Survey; Voges et al. 1999), from which a number of cluster catalogues were built and used for cosmology (e.g., Ebeling et al. 1998; Böhringer et al. 2000, 2004).

Optical cluster catalogues have also improved significantly over the last decades, and have been used for cosmology since the 2000s. A turning point was the Sloan Digital Sky Survey (SDSS), from which catalogues with over 10 000 clusters have been produced and used in cosmological analyses (e.g., Koester et al. 2007; Costanzi et al. 2019).

The most recent addition to the rich corpus of galaxy cluster surveys and catalogues are those based on the tSZ effect, whose first robust detection was reported in Birkinshaw et al. (1984). Over the last decade, the CMB experiments *Planck*, ACT, and SPT (and the subsequent incarnations of the two latter experiments) have produced a number of SZ cluster catalogues with up to around 1 000 clusters in them (e.g., Staniszewski et al. 2009; Vanderlinde et al. 2010; Sehgal et al. 2011; Bleem et al. 2015; Planck 2015 results XXVII 2016; Hilton et al. 2018). In these catalogues, clusters are blindly detected from CMB frequency maps thanks to the characteristic tSZ spectral signature, typically by using matched filters (e.g., Herranz et al. 2002; Melin et al. 2006).

The cosmological constraints obtained from these X-ray, optical, and SZ surveys over the past few years are discussed in Section 1.5.

1.1.3 Inferring information from observations

Cosmologists (and, more generally, scientists) often have a model with which to describe their observations that depends on a set of parameters whose values they want to determine. This will be the case in Chapter 3, in which we constrain the values of several cosmological and astrophysical parameters with galaxy cluster data, and in Chapter 4, in which we constrain the values of several cluster scaling relation parameters with simulations. For illustration, in cluster cosmology, a parameter of interest can be, for example, $\Omega_{\rm m}$, which parametrises the mean matter density of the Universe (see Section 1.2.1).

The values of the parameters of interest are typically determined (or inferred) by 'fitting the model to the data'. The rigorous way to do this is by using probability theory (e.g., Jaynes 2003 or Mackay 2003). In particular, one has to be able to compute the *likelihood*, which is the probability of obtaining the measured data given a set of parameter values. Denoting our data with D and our set of model parameters with θ, the likelihood is thus the probability of D conditioned on θ, $P(D|\theta)$. The quantity of interest, however, is the probability of the parameters given the data, $P(\theta|D)$, known as the *posterior*. If the likelihood is known, the posterior can be easily computed using Bayes' theorem[3],

$$P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)}. \tag{1.1}$$

Here, $P(\theta)$, commonly referred to as the *prior*, is the unconditioned probability of the parameters, and $P(D)$, known as the *evidence*, is the unconditioned probability of the data, which does not depend on the parameters and can be obtained by normalising $P(\theta|D)$ to 1. The prior, $P(\theta)$, can be seen as encapsulating our knowledge of the parameters before the data D is taken into account (more precisely, the prior knowledge that we want to consider). Multiplying it by the likelihood in order to obtain the posterior can be seen as 'updating' this prior knowledge with the information provided by the data.

[3]This theorem, which is simply a rearrangement of the product rule of probability, currently gives its name to the whole practice of posterior-based inference. As noted in Jaynes (2003), this result had already been derived by others before Bayes, such as Jacob Bernoulli and de Moivre, and it was Laplace (Laplace, 1774) who recognised its power as a tool for inference.

The product of this inference practice is the posterior, $P(\theta|D)$, which has to be 'explored'. If the parameter space is low-dimensional, simple evaluation of the posterior over a grid in parameter space can be a quick and, especially, simple way to do this. For instance, if there is just one parameter, this procedure will yield a curve that is the posterior probability density function of this only parameter. However, this becomes computationally expensive as the dimension of the parameter space increases, since its time complexity grows exponentially with the number of dimensions. A widely-used alternative more suitable for higher dimensional parameter spaces are the family of Markov chain Monte Carlo (MCMC) methods (see, e.g., Mackay 2003). There are a number of different algorithms that fall within this category, but their basic idea is the same: in order to explore the posterior, one, or several, random chains explore the parameter space. After some *burn-in* chain iterations, their positions in parameter space correspond to (correlated) samples from the posterior. These samples can then be used to visualise the posterior, e.g., by producing one-dimensional histograms (or, alternatively, density estimates) for each parameter, which correspond to the marginalised posteriors of each parameter, or two-dimensional histograms (or density estimates) for each possible pair of parameters, which will reveal correlations (or degeneracies) between them. They can also be used in order to estimate the mean and median values of each parameter, confidence intervals (e.g., by computing some quantiles), and any other statistic of interest, which may be used to summarise the parameter constraints.

1.1.4 Cosmology with galaxy clusters

As mentioned in Section 1.1.1, there are two main ways in which cosmological information can be extracted from galaxy clusters: by looking at their abundance as a function of mass and redshift (the 'cluster counts' analyses), and by looking a the gas mass relative to the total cluster mass (the 'gas fraction' analyses). In this section we review in a general way the principles on which they are based, and also briefly discuss other interesting avenues to obtain cosmological information from clusters.

Galaxy cluster counts

The two fundamental quantities characterising a galaxy cluster are its mass[4], M, and its redshift[5], z. The abundance of clusters as a function of these two quantities per unit

[4]In this section we refer to mass in a generic way; precise definitions of cluster mass used in practice are given in Section 1.2.4.

[5]For a definition of redshift, see Section 1.2.1

solid angle Ω, i.e., $\frac{d^4N}{dMdzd^2\Omega}(M,z)$, where $\frac{d^4N}{dMdzd^2\Omega}(M,z)dMdzd^2\Omega$ is the mean number of clusters within an infinitesimal volume $dMdzd^2\Omega$ around the point (M,z,Ω), depends on the assumed cosmological model and on the values of its parameters. This mean abundance can be estimated from observations, and it can also be predicted, for which cosmological simulations are typically used (see Section 1.2.5). Therefore, following the ideas outlined in Section 1.1.3, it should be possible to obtain cosmological information from the observed cluster abundance, e.g., to constrain the values of some cosmological parameters assuming a cosmological model. This is the basic idea behind cluster counts. The precise cosmological parameters that cluster counts are most sensitive to are discussed in Section 1.2.5.

Redshifts can be observed directly without much difficulty. The determination of cluster masses, however, is a complicated art, as we will see in Section 1.3; currently, mass determination is the main source of systematic uncertainty in cluster counts studies (Pratt et al., 2019). Instead of mass estimates, cluster observables that are known to scale with cluster mass, e.g., the X-ray luminosity or the SZ amplitude, are often used, since they are easier to obtain observationally. These are often referred to as 'mass proxies' or 'mass observables' (Allen et al., 2011).

A typical cluster counts data set is a catalogue of clusters with their redshift and their mass proxy (or estimate) values. Several mass proxies can be used for each cluster, and it is not necessary that all the clusters have a measured value for each of the proxies used in the analysis. In order for this catalogue to be used for cosmology, the distribution of observed cluster counts across the cluster observable space (redshift–mass proxy(s) space) has to be predicted as a function of the cosmological parameters of interest. If correlations between different clusters can be neglected, a possible way to approach this is to divide the cluster observable space into n bins. The data then simply reduces to number counts across the bins, and the counts within each bin follow a Poisson distribution. Thus, the likelihood is just a product of Poisson distributions, one for each bin, and the cosmology-dependent quantity to be predicted is the expected number of detected clusters within each bin. This was the approach followed, e.g., in Planck 2015 results XXIV (2016). An alternative approach in which no binning is used is followed in Chapter 3; how it relates to the traditional binning approach is detailed in Appendix C. If correlations between clusters cannot be ignored, the covariance due to their spatial clustering has to be taken into account; see, e.g., Hu and Kravtsov (2003) and Smith and Marian (2011) for how this can be done.

For any of these approaches to work, several ingredients have to be taken into account. First, the cluster mass proxy(s) must be related to the cluster mass (or, if they

are mass estimates, they must be unbiased). Second, the selection of the sample has to be taken into account. For example, in the binned Poisson approach, the expected number of observed clusters within each bin must be the expected number of observed clusters within that bin given the selection criteria followed to construct the cluster sample used. Finally, this selection needs to be well understood. In particular, the probability of selecting an observed cluster (i.e., of including it in the sample) at given mass and redshift, sometimes referred to as the *completeness* (Allen et al., 2011), must be a well-determined quantity. In addition, the catalogue must have a high *purity*, i.e., it must contain as few false detections as possible.

Relating mass proxies (or mass estimates) to cluster mass is done through scaling relations, which assign a mean value of the mass proxy to each point in cluster mass–redshift space, which is the space in which the cluster abundance is robustly predicted. These scaling relations often are parametrised as power laws of the mass, and are calibrated with simulations and/or with careful estimates of cluster masses, the latter typically for a subsample of the cluster sample considered in the analysis (as in, e.g., Planck 2015 results XXIV 2016).

There is always scatter around the mean proxy values given by the scaling relations, and it has to be well understood. This scatter can be of two types: 'observational scatter' (or observational 'noise'), and 'intrinsic scatter'. The former refers to the uncertainty in the mass proxy measurements due to observational limitations (e.g., due to the finite resolution of the telescope used for the observations, to instrumental noise, etc.). It can be reduced with better observations of the same clusters. On the other hand, intrinsic scatter refers to the *intrinsic* variability in the cluster observables at given mass and redshift due to the clusters themselves, and to their surroundings. An example of a source of intrinsic scatter is triaxiality. Clusters are in general triaxial: at a given redshift, two clusters with the same mass have, in general, different shapes and orientations, and this translates into them having different cluster observable values (e.g., X-ray luminosities). For a given observable, intrinsic scatter is irreducible: it cannot be beaten down with better observations. However, some observables have larger intrinsic scatter than others: in order to make the most of the data, observables with small intrinsic scatter must be used. In general, a good mass proxy must have little total scatter (observational plus intrinsic), depend sensitively on cluster mass, have a well-determined scaling relation, and be easy to measure (Allen et al., 2011).

As it has already been noted, more than one mass proxy and/or estimate can be used in the same counts analysis, and indeed this can prove a very powerful idea, as different proxies and/or estimates can be very complimentary. For example, some have small

scatter, but their mean scaling relation may be subject to poorly-understood biases (e.g., X-ray proxies), whereas others have larger scatter, but their mean dependence on cluster mass is well understood (e.g., lensing mass estimates). Using them jointly in the same analysis allows the noisy, unbiased proxy(s) (or estimate(s)) to calibrate the mean scaling relation of the high signal-to-noise proxy(s) (or estimate(s)), yielding both precise and unbiased cosmological constraints. This is the idea behind the counts analysis presented in Chapter 3, in which very noisy but nearly unbiased CMB lensing cluster mass estimates are used in conjunction with much higher signal-to-noise SZ mass proxies, which have a mean scaling relation in which there is an unknown mean mass bias parameter, the so-called hydrostatic mass bias. As highlighted in Chapter 3, if several mass proxies are used in an analysis, it is crucial that they are included self-consistently in the likelihood in order to avoid biases in the results (see, e.g., Appendix B), and their intrinsic correlations must be taken into account.

Gas fraction

The basic idea behind gas fraction analyses is that, given the huge sizes of galaxy clusters, the fraction of cluster mass that is in the form of gas (which, as noted in Section 1.1.2, dominates the cluster baryonic mass) is approximately equal to the mean universal ratio of baryonic matter to total matter. The small deviations due to, e.g., star formation, can be corrected for with simulations. In particular, the gas fraction, $f_{\rm gas}$, can be written as (Allen et al., 2011)

$$f_{\rm gas} = \frac{M_{\rm gas}}{M} = Y(z)\frac{\Omega_{\rm b}}{\Omega_{\rm m}}, \tag{1.2}$$

where $M_{\rm gas}$ is the cluster gas mass (which can be measured with, e.g., X-ray observations); M is the cluster total mass; $Y(z)$ is the depletion parameter, which quantifies deviations from the universal ratio and which can be determined with simulations; and $\Omega_{\rm b}$ and $\Omega_{\rm m}$ parametrise, respectively, the present-time mean baryonic and total matter density of the Universe (see Section 1.2.1).

If $\Omega_{\rm b}$ is determined from other observations (e.g., with observations of the CMB anisotropies, or of the relative abundances of light elements in the Universe), gas fraction measurements allow to constrain the value of $\Omega_{\rm m}$. This provided the first piece of compelling evidence that $\Omega_{\rm m} < 1$ (White et al., 1993).

In addition, the measured values of $f_{\rm gas}$ depend on the assumed distances to the clusters used in the analysis, since they are obtained from integrated quantities. In turn, these distances, as noted in Section 1.2.1, depend on cosmology. This brings in

additional sensitivity to cosmology, in particular to the nature of dark energy (Sasaki, 1996). Following this idea, Allen et al. (2004) provided the first direct measurement of the acceleration of the Universe independent from the supernovae Ia measurements that discovered it.

Other probes

Although cluster counts and the gas fraction are the most widely used ways to obtain cosmological information from galaxy clusters, other methods exist.

One of them, first proposed in Silk and White (1978), is based in combining X-ray and SZ measurements of the same galaxy clusters. The idea is that the SZ signal for each cluster can be predicted with X-ray observations; making the predicted SZ signal match the observed signal brings in sensitivity to cosmology via the distances involved in the prediction.

Another possible probe is the angular power spectrum of the tSZ signal across the sky, as first proposed in Komatsu and Seljak (2002). This power spectrum can be measured with CMB observations, and can be predicted assuming models of cluster abundance, the cluster SZ signal profile, and the spatial clustering of clusters, the latter being only a small correction. This probe is particularly sensitive to σ_8, which parametrises the amplitude of the large-scale matter fluctuations in the Universe (see Section 1.2.3). The kinetic SZ cluster signal can also be exploited in order to obtain cosmological constraints (see Section 1.3.5).

The most massive galaxy cluster found in a given survey (or in several surveys) can also be used in order to constrain cosmological models: if the cluster abundance across mass and redshift can be predicted, the probability of the highest mass in the sample can be calculated. This is particularly useful to test the validity of a given cosmological model, as the existence of an extremely unlikely very massive cluster within that model would allow to rule it out (see, e.g., Mortonson et al. 2011). Parameter constraints assuming a given cosmological model can also be obtained (e.g., Sahlén et al. 2016).

Finally, galaxy clusters can also be used to investigate the physical nature of dark matter. An interesting 'laboratory' in this respect is the Bullet Cluster, which in fact consists of two merging clusters, one of which has just gone through the other one. A composite image of the Bullet Cluster is shown in Figure 1.5, in which the X-ray emitting gas is shown in pink, and the mass distribution from a lensing analysis is shown in blue. It can be observed how, in each of the two clusters, the galaxies and the dark matter, which accounts for most of the lensing signal, have roughly the same centres, which means that they have undergone little interaction. The two clusters'

Fig. 1.5 Bullet cluster, two merging clusters at $z = 0.52$. The X-ray emission from the ICM gas is shown in pink, whereas the matter distribution from lensing is shown in blue. Figure credit: `chandra.harvard.edu`; Markevitch (2006) (X-ray), Clowe et al. (2006) (optical and lensing).

ICMs, however, have undergone significant interaction, being significantly away from their corresponding galaxies and dark matter. The smaller ICM, on the right, also features a characteristic bow shock pattern. This apparent absence of dark matter interaction allows to put upper bounds on the dark matter self-interaction cross-section (e.g., Markevitch et al. 2004; Randall et al. 2008).

1.2 From quantum fluctuations to galaxy clusters

In this section we review, from a theoretical point of view, how galaxy clusters form in our expanding Universe, and what we know about their abundance, their spatial distribution, and their internal structure.

1.2.1 The expanding Universe

The dynamics of the Universe

Assuming that the Universe is homogeneous and isotropic on very large scales, on these scales spacetime can be described with the Friedmann-Lemaître-Robertson-Walker (FLRW) metric (Friedmann, 1922; Lemaître, 1927; Robertson, 1932; Walker, 1935),

$$ds^2 = -c^2dt^2 + a(t)^2d\mathbf{x}^2, \tag{1.3}$$

where t is a time coordinate, known as cosmic time, $a(t)$ is the scale factor, which only depends on t, and $\mathbf{x}$ are known as the comoving spatial coordinates, which are fixed for observers whose only relative motion is the large-scale expansion (or contraction) of the Universe (comoving observer). Here, $d\mathbf{x}^2$ describes the comoving spatial part of the metric, which has constant curvature, k. This curvature can be positive (closed universe), negative (open universe), or zero (flat universe). The time evolution of the Universe is thus just given by the evolution of the scale factor, a, which we normalise to 1 at the present time, t_0, i.e., $a(t_0) = 1$.

An important concept in order to link the evolution of the scale factor with observations is that of redshift, z. This is the wavelength (or frequency) shift towards larger wavelengths that radiation undergoes when emitted by distant sources (e.g., galaxies) due to the expansion of the Universe. Defining it as $z \equiv (\lambda_r - \lambda_e)/\lambda_e$, where λ_r and λ_e are, respectively, the received and emitted wavelengths, it can be shown (e.g., Dodelson 2004) that

$$1 + z = \frac{1}{a}, \tag{1.4}$$

where a denotes the scale factor at the time of emission. Thus, z and a can be used interchangeably to parametrise the expansion of the Universe.

In order to study the time evolution of the scale factor, the metric must be related to the distribution of energy and momentum in the Universe. Assuming general relativity, this is done through the Einstein's equations (Einstein, 1915),

$$G_{\mu\nu} + \Lambda g_{\mu\nu} = \frac{8\pi G}{c^4} T_{\mu\nu}, \tag{1.5}$$

where $G_{\mu\nu}$ is the Einstein tensor, which is given by the metric, $g_{\mu\nu}$, and its first two derivatives, Λ is the cosmological constant, and $T_{\mu\nu}$ is the energy-momentum tensor. As a consequence of homogeneity and isotropy, $T_{\mu\nu}$ reduces to a density, ρ, and an isotropic pressure, p, which are only functions of time.

Imposing Einstein's equations to the FLRW metric and to this energy-momentum tensor yields the Friedmann equations,

$$H^2 = \frac{8\pi G}{3}\rho - \frac{kc^2}{a^2} + \frac{\Lambda c^2}{3}, \tag{1.6}$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}\left(\rho + \frac{3p}{c^2}\right) + \frac{\Lambda c^2}{3}, \tag{1.7}$$

where $H \equiv \dot{a}/a$ is known as the Hubble parameter, and where overdots denote derivatives with respect to cosmic time, t.

Combining the two Friedmann equations yields the continuity equation,

$$\dot{\rho} + 3H\left(\rho + \frac{p}{c^2}\right) = 0, \tag{1.8}$$

which expresses energy conservation.

In order to solve the Friedmann equations, pressure, p, must be related to density, ρ, through an equation of state, $p(\rho)$. Different energy components have different equations of state. Our Universe can be approximated as a mixture of non-relativistic matter and radiation. For non-relativistic matter, the density largely dominates over the pressure, so we can set $p = 0$. For radiation, $p = \rho c^2/3$. Furthermore, the cosmological constant term in the Friedmann equations can be seen as arising from an additional component, (constant) dark energy, with constant density $\rho_\Lambda = \Lambda c^2/8\pi G$, and with pressure $p = -\rho_\Lambda c^2$. Note that it is possible to consider other forms of dark energy in which $w \neq -1$, where $w = p/(\rho c^2)$; all current observations are compatible with $w = -1$ (Huterer and Shafer, 2018).

Assuming that there are no interactions between the different components, their scalings with a can be derived from the continuity equation. In particular, for matter, $\rho \propto a^{-3}$, and for radiation, $\rho \propto a^{-4}$. Assuming the Universe to be a mixture of radiation, matter, and constant dark energy, the first Friedmann equation can be written as

$$H^2 = \frac{8\pi G}{3}\left(\frac{\rho_{\mathrm{r},0}}{a^4} + \frac{\rho_{\mathrm{m},0}}{a^3} + \rho_\Lambda\right) - \frac{kc^2}{a^2}, \tag{1.9}$$

where $\rho_{\mathrm{r},0}$ and $\rho_{\mathrm{m},0}$ are, respectively, the present-time radiation and matter densities. Defining the critical density as the total density for which the Universe is flat ($k = 0$), $\rho_{\mathrm{c}} = 3H^2/8\pi G$, we can define density parameters as the ratio of the density of each component to the critical density. The first Friedmann equation can then be rewritten as

$$H^2 = H_0^2\left(\frac{\Omega_{\mathrm{r}}}{a^4} + \frac{\Omega_{\mathrm{m}}}{a^3} + \frac{\Omega_k}{a^2} + \Omega_\Lambda\right) = H_0^2\left[\Omega_{\mathrm{r}}(1+z)^4 + \Omega_{\mathrm{m}}(1+z)^3 + \Omega_k(1+z)^2 + \Omega_\Lambda\right], \tag{1.10}$$

where H_0 is the present-time Hubble parameter (the 'Hubble constant'), Ω_r, Ω_m, and Ω_Λ are, respectively, the present-time density parameters of radiation, matter, and dark energy, and where a density parameter for the curvature term has also been defined, $\Omega_k = 1 - \Omega_\mathrm{r} - \Omega_\mathrm{m} - \Omega_\Lambda$. The Hubble constant and the density parameters are some of the most important cosmological parameters, since they quantify, respectively, the Universe's present-time expansion rate and the energy densities of each of the components of the Universe. Furthermore, together they determine the expansion history of the Universe. Another important parameter is Ω_b, which is the present-time density parameter of *baryonic* matter. If we denote the present-time density parameter of dark matter with Ω_c, then $\Omega_\mathrm{b} + \Omega_\mathrm{c} = \Omega_\mathrm{m}$.

Current CMB observations, in combination with baryon acoustic oscillation (BAO) measurements, yield, assuming spatial flatness, $h = 0.6766 \pm 0.0042$, $\Omega_\mathrm{m} = 0.3111 \pm 0.0056$, $\Omega_\Lambda = 0.6889 \pm 0.0056$, and $\Omega_\mathrm{b}h^2 = 0.02242 \pm 0.00014$ ($1\,\sigma$ confidence limits; Planck 2018 results VI 2018), where h is the Hubble parameter in units of $100\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$. Allowing non-zero spatial curvature yields $\Omega_k = 0.0007 \pm 0.0019$ (Planck 2018 results VI, 2018). In addition, measurements of the CMB mean absolute temperature, $T_\mathrm{CMB} = 2.7255 \pm 0.0006\,\mathrm{K}$, imply $\Omega_\mathrm{r}h^2 = 2.47 \times 10^{-5}$ (Mather et al., 1994; Fixsen, 2009). These measurements are in agreement with a number of other independent measurements, except, perhaps, that of H_0, with current direct measurements finding somewhat higher values (e.g., Riess et al. 2019; see Knox and Millea 2019).

Thus, at present time, the energy density of the Universe is dominated by dark energy (about 70 %), with also a significant contribution from matter (about 30 %), of which only about 20 % is baryonic. In addition, the Universe is consistent with being spatially flat. However, in the past the situation was different, due to the different scaling of each component with redshift. Indeed, at very early times, the Universe was radiation dominated. At about $z \simeq 3400$, it transitioned to be matter dominated, and only at $z \simeq 0.3$ did it become dark energy dominated. Also, extrapolating this model to the distant future, the Universe will become completely dominated by dark energy, which implies that the scale factor will grow exponentially with time.

Distances in cosmology

Distance is an important concept in cosmology. Assuming the FLRW metric (see Eq. 1.3), the *comoving* distance between two points at a given time t, χ, is defined by integrating the comoving coordinate line element, $d\mathbf{x}$, along the (spatial) geodesic connecting them. If two objects are comoving (i.e., their only relative motion is the Hubble flow), the comoving distance between them remains constant with time. If

$k = 0$, the comoving distance simply reduces to the Euclidean distance between the two points in comoving coordinates. Multiplying χ by the scale factor gives the *proper* distance, $d_\mathrm{p} = a\chi$, which is the physical spatial distance between two points at a given time as would be measured by a comoving observer. Since at the present time $a = 1$, the comoving distance between two comoving points corresponds to their present-time physical distance.

Despite being conceptually useful, these two distances cannot be measured directly. Two directly measurable distances can be defined: the *luminosity* distance, and the *angular diameter* distance. The luminosity distance, d_L, is defined as the distance to an object that is inferred from its observed flux, F, i.e.,

$$F = \frac{L}{4\pi d_\mathrm{L}^2}, \tag{1.11}$$

where L is the object's luminosity.

On the other hand, the angular diameter distance, d_A, is defined as the distance to an object that is inferred from its apparent size, θ, i.e.,

$$\theta = \frac{l}{d_\mathrm{A}}, \tag{1.12}$$

where l is the object's physical length.

For a spatially-flat universe, it can be shown (see, e.g., Peacock 1999) that $d_\mathrm{L} = d_\mathrm{p}(1+z)$, and $d_\mathrm{A} = d_\mathrm{p}/(1+z)$, where z is the object's redshift. (Note, however, that $d_\mathrm{L} = d_\mathrm{A}(1+z)^2$ is always valid, regardless of curvature.)

The comoving distance to a source (e.g., a galaxy), and therefore the luminosity and angular diameter distances, can be expressed as a function of redshift, z, by integrating the FLRW metric for an incoming photon. The resulting expression depends on the expansion history of the Universe, and therefore on the values of H_0 and of the density parameters.

1.2.2 The seeds of cosmic structure

Thus far we have only considered a perfectly homogeneous and isotropic Universe, which is a good description on the very largest scales. However, on smaller, but still cosmic, scales inhomogeneities (and, indeed, anisotropies) exist, from stars to the cosmic web. A useful quantity with which to characterise these inhomogeneities is the power spectrum (equivalently, the 2-point correlation function in real space). Imposing, due to large scale homogeneity and isotropy, that statistical quantities cannot depend

on position or direction, the power spectrum of a given field X (e.g., the matter density perturbation) takes the following form (e.g., Dodelson 2004):

$$\langle X(\mathbf{k})X^*(\mathbf{k}')\rangle = \delta^{(3)}(\mathbf{k}-\mathbf{k}')P_X(k). \tag{1.13}$$

That is, modes with different wavelengths are uncorrelated, and the power spectrum only depends on the modulus of the wavevector. It is also useful to define the dimensionless power spectrum, $\Delta^2_X(k) \equiv P_X(k)k^3/(2\pi^2)$, which can be seen as the contribution to the total field variance per logarithmic interval as a function of scale k, since the total variance is just the integral of $P(k)$ over all $\mathbf{k}$-space.

The evolution of inhomogeneities as the Universe expands is well understood in the linear regime (i.e., when they are small), in which they can be computed analytically (see Section 1.2.3), and, to a lesser extent, in the nonlinear regime, thanks to some analytical insights and to numerical simulations (see Section 1.2.4 and subsequent sections). The origin of cosmic inhomogeneities is, however, less well-established. The currently-followed paradigm is that of cosmic inflation, according to which they are set up from quantum fluctuations very shortly after the Big Bang, and then they 'passively' evolve under the action of the laws of physics (on large scales, mostly gravity) to produce the rich hierarchy of structure that we observe today. The highest, rarest peaks of this primordial density field evolve to become galaxy clusters at late times.

Cosmic inflation, originally proposed in the early 1980s (Guth, 1981; Linde, 1982, 1983), invokes an 'inflaton' field (or several, even many, of them[6]) which dominates the energy content of the Universe at very early times, before the Standard Model particles are produced, and which makes the Universe expand in an accelerated way. The key feature of accelerated expansion is that the Hubble horizon, c/H, roughly the size of the regions of the Universe that are causally connected within a Hubble time, $1/H$, grows less quickly than the scale factor. Equivalently, the comoving Hubble horizon, $c/(aH)$, decreases with time. This is the opposite to what happens in decelerated expansion, in which the comoving Hubble horizon grows with time (or, equivalently, the physical Hubble horizon grows more quickly than the scale factor). This means that regions that were originally causally connected within a typical expansion time cease to be so as inflation proceeds. This can be seen graphically in Figure 1.6, in which the comoving Hubble horizon is shown, schematically, as a function of the scale factor during and after inflation. For a given comoving scale λ, the relevant scale is that of 'horizon crossing', at which $\lambda \sim c/(aH)$. When inflation ends and the Universe starts its 'standard', decelerated expansion (until dark energy causes a new

[6]For concision, in the following we only refer to one field.

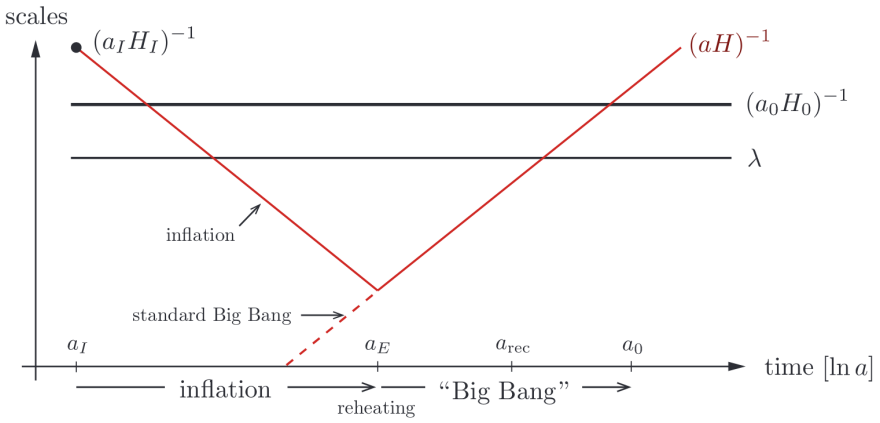


Fig. 1.6 Schematic view of inflation. The comoving Hubble horizon is shown in red (note that $c = 1$). It can be seen how it decreases during inflation, and how it increases again during standard Big Bang evolution. The present-time comoving Hubble horizon, $(a_0 H_0)^{-1}$, and a reference comoving scale currently within the Hubble horizon, λ, are shown as horizontal lines. Figure credit: Daniel Baumann's Part III of the Mathematical Tripos Cosmology lecture notes; `theory.uchicago.edu/~liantaow/my-teaching/dark-matter-472/lectures.pdf`.

period of accelerated expansion), the comoving Hubble horizon starts growing again, progressively allowing regions to be causally connected again as they enter back into the horizon (see Figure 1.6).

This accelerated expansion solves several problems of the standard cosmological paradigm, which was the reason why inflation was originally proposed (e.g., it solves the horizon problem and the flatness problem; see, e.g., Dodelson 2004). Inflation, however, also provides a mechanism to set up the primordial fluctuations (e.g., Mukhanov and Chibisov 1981). The basic idea is the following. Present-time cosmological scales corresponded to very small, quantum scales at inflationary times. The field driving inflation has quantum zero-point fluctuations, and since it dominates the energy budget of the Universe during inflation, these fluctuations translate into perturbations in the metric. These perturbations become classical, stochastic fluctuations when they are stretched out to classical scales by the expansion, which, in turn, translate into perturbations in the energy-momentum of the Standard Model (plus dark matter) fields that come to dominate the Universe after inflation ends.

Current data is consistent with these primordial perturbations being Gaussian (i.e., fully described by their power spectrum), scalar, adiabatic, and almost scale-invariant (e.g., Planck 2018 results X 2018). Scalar perturbations are one of the three possible types of perturbations into which fluctuations can be decomposed, the other two

types being vector and tensor perturbations. Vector perturbations are not expected to be produced by inflation, and decay quickly with the expansion anyway, leaving no observational signatures. Tensor perturbations, however, are generically predicted by inflation models and can leave potentially detectable observational signatures (e.g., B-modes in the CMB polarisation pattern, see Section 1.4), which are yet to be found observationally (e.g., BICEP2 Collaboration and Keck Array Collaboration 2018). Adiabatic perturbations are one of the two possible types of scalar perturbations (the other being isocurvature perturbations), in which the perturbations in the different components (e.g., radiation and dark matter) are associated to the same perturbation in the curvature. They can be seen as produced by locally compressing or expanding all the components in the Universe at once. If there are only adiabatic scalar perturbations, the primordial perturbations can be fully described with the so-called comoving curvature perturbation, $\mathscr{R}(\mathbf{k})$, which is conserved on scales larger than the Hubble horizon (Bardeen, 1980). If perturbations are Gaussian, $\mathscr{R}(\mathbf{k})$ in turn is fully described by its power spectrum, which is often parametrised with a power law. The associated dimensionless power spectrum can be written as

$$\Delta^2_{\mathscr{R}}(k) = A_s \left(\frac{k}{k_\star}\right)^{n_s - 1}, \tag{1.14}$$

where $k_\star$ is a pivot scale, typically chosen to be $0.05\,\mathrm{Mpc}^{-1}$ (e.g., Planck 2018 results X 2018), A_s is the amplitude at the pivot scale, and n_s is known as the scalar spectral index. If $n_s = 1$, the power spectrum is said to be scale-invariant, since there is the same amount of variance in the curvature per logarithmic interval of k: the Universe 'looks the same' on all scales. Current observations yield $\ln(10^{10} A_s) = 3.047 \pm 0.014$ and $n_s = 0.9665 \pm 0.0038$ (Planck 2018 results VI, 2018), which means that the primordial perturbations are almost scale-invariant.

Inflation fits well within this scenario. For the simplest inflationary models, those with a scalar field slowly rolling down a potential, $\Delta^2_{\mathscr{R}}(k)$ well after horizon crossing can be written as (Baumann, 2009)

$$\Delta^2_{\mathscr{R}}(k) = \frac{1}{8\pi^2} \frac{1}{\epsilon} \frac{H^2}{M^2_{\mathrm{Pl}}}, \tag{1.15}$$

where ϵ is the first slow-roll parameter, which, during inflation, $\epsilon \ll 1$, and M_{Pl} is the Planck mass. Here, H and ϵ are evaluated at horizon crossing for each wavelength k, since, upon horizon exit, a curvature perturbation mode 'freezes', only to evolve again when it enters back into the horizon after inflation has finished. The fact that the

spectrum is almost scale invariant is a consequence of H and ϵ being approximately constant as inflation proceeds.

There is currently a large number of inflationary models that can generate these Gaussian, almost scale-invariant, adiabatic, scalar perturbations (Planck 2018 results X, 2018): the physical nature of the inflaton is thus very uncertain. A detection of, e.g., deviations from Gaussianity (e.g., by detecting higher-order correlations), or of primordial tensor modes (e.g., by detecting primordial B-modes in the CMB polarisation), would provide valuable evidence that inflation actually happened and would cast light on the physics responsible for it (Bartolo et al., 2004).

1.2.3 Growing structure: linear evolution

Linear growth of structure

The evolution of perturbations with time is well understood at the linear level (e.g., Lifshitz 1946; Peebles and Yu 1970; Peebles 1980; Bardeen 1980; Kodama and Sasaki 1984; Mukhanov et al. 1992; Seljak et al. 2003). An important property of linear evolution is that it preserves the Gaussian nature of the perturbations, with different modes evolving independently; 'true', nonlinear evolution does not. For our purposes, the relevant quantity to track is the matter perturbation, or matter density contrast, δ, which is defined as

$$\delta(\mathbf{x}, a) \equiv \frac{\rho_{\rm m}(\mathbf{x}, a) - \bar{\rho}_{\rm m}(a)}{\bar{\rho}_{\rm m}(a)}, \tag{1.16}$$

where $\rho_{\rm m}(\mathbf{x}, z)$ is the matter density at location $\mathbf{x}$ and scale factor a, and $\bar{\rho}_{\rm m}(z)$, the mean matter density at redshift z. After recombination, its power spectrum as a function of redshift can be related to the primordial comoving curvature power spectrum, $P_{\mathscr{R}}(k)$, as

$$P_{\rm m}(k, a) = G^2(a) T^2(k) P_{\mathscr{R}}(k). \tag{1.17}$$

Here, $G(a)$ is the linear growth function, which encapsulates the growth of matter perturbations due to gravity after recombination. In the current ΛCDM paradigm, perturbations grow independently of their scale, i.e., the growth function does not depend on k. During matter domination, δ grows approximately as a, whereas during Λ domination its growth stalls, remaining constant, due to the rapid, accelerated expansion (e.g., Dodelson 2004). In general, in ΛCDM, $G(a)$ can be written as (Heath, 1977)

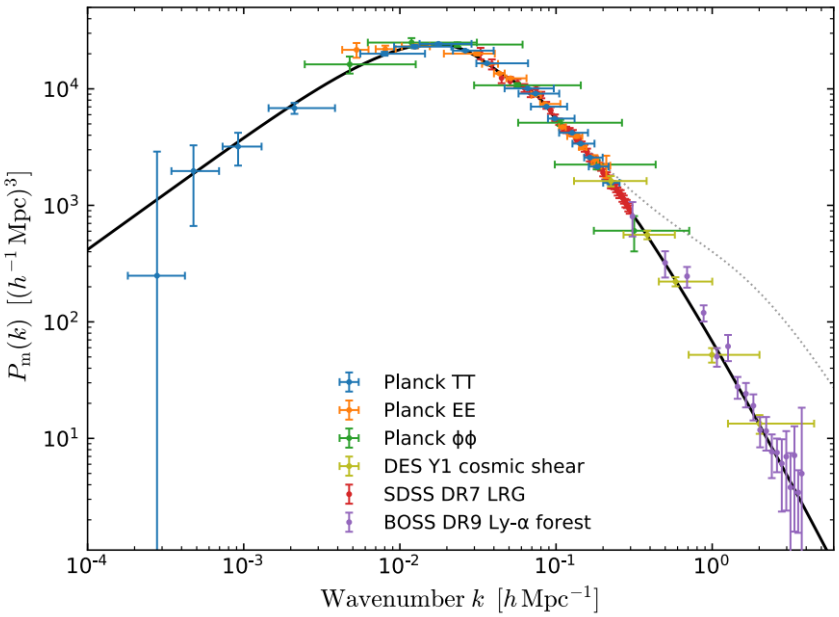


Fig. 1.7 Linear-theory matter power spectrum at $z = 0$ inferred from different cosmological probes. The black solid line is the prediction of ΛCDM with the *Planck* best-fit parameters, showing remarkable consistency, and the grey dotted line is the nonlinear prediction. Figure credit: Planck 2018 results I (2018).

$$G(a) = \frac{5}{2}\Omega_{\mathrm{m}} E(a) \int_0^a \frac{da'}{(a'E(a'))^3}, \tag{1.18}$$

where $E(a) \equiv H(a)/H_0$, and where the normalisation is such that, at high redshift, in the matter domination epoch, $G(a) \propto a$.

On the other hand, $T(k)$ is the transfer function, which describes evolution before recombination, which is markedly k-dependent. The main contribution to this k-dependence is that the growth of perturbations due to gravity is slowed down for modes that enter the horizon during radiation domination due to mostly radiation-sourced pressure, which competes against gravity (Meszaros, 1974). This can be seen in Figure 1.7, which shows the present-time matter power spectrum, $P_{\mathrm{m}}(k)$. On large scales, which entered the horizon well after matter-radiation equality, $P_{\mathrm{m}}(k) \propto k^{n_s}$ (approximately), whereas, on scales that entered the horizon well before equality, $P_{\mathrm{m}}(k) \propto k^{n_s-4}$ (also approximately).

Another important scale-dependent effect is baryonic acoustic oscillations (BAOs) (Holtzman, 1989). During radiation domination, sound waves propagate in the tightly coupled baryon-radiation fluid due to the mostly radiation-sourced pressure. These sound waves also affect the total matter distribution via gravity. They have an

associated length scale, the sound horizon, which is the distance that they have travelled since the Big Bang until a given time. This scale freezes at decoupling, when photons, the main source of pressure, decouple from baryons to free-stream ever since, and as a result the sound horizon at decoupling is imprinted in the matter distribution (and also in the photon distribution, as we will see in Section 1.4). It manifests as a local peak in the otherwise decreasing-with-distance matter 2-point correlation function: two points separated by a distance equal to the sound horizon at decoupling are more correlated than they would be if they were separated by slightly shorter or longer distances. In Fourier space, this local peak translates into oscillations in the power spectrum on small scales. These oscillations can be seen very clearly in the power spectrum of the CMB, as we will see in Section 1.4.

Filtering

An important concept in the theoretical understanding of the galaxy cluster population is that of filtering. Consider the matter density contrast field $\delta(\mathbf{x}, a)$. We can filter it with a window function to produce a filtered density contrast, δ_R, as

$$\delta_R(\mathbf{x}, a) = \int \delta(\mathbf{x}', a) W_R(|\mathbf{x} - \mathbf{x}'|) d^3\mathbf{x}', \tag{1.19}$$

where W_R is a window function, or filter, with a characteristic scale R; typical choices are real-space top-hat, Gaussian, or Fourier-space top-hat. The standard deviation of the filtered field, $\sigma(R, a)$, can be written as (e.g., Zentner 2007)

$$\sigma(R, a) = \frac{1}{(2\pi)^{3/2}} \left(\int P_{\rm m}(k, a) |W(\mathbf{k}, R)|^2 d^3\mathbf{k} \right)^{1/2}, \tag{1.20}$$

where $P_{\rm m}(k, a)$ is the matter power spectrum, and $W(\mathbf{k}, R)$ is the Fourier transform of the window function. After recombination, $\sigma(R, a)$ grows as δ, i.e., with the growth function. At a given time, it is possible to associate a mass to each filter scale; for a top-hat filter, this is just $M = 4/3\pi\bar{\rho}_{\rm m}(a)(aR)^3$. Thus, the filtered matter density contrast field, and its standard deviation, can be parametrised with either R or M.

The standard deviation of the present-time linear density field filtered with a real space top-hat window function with scale $R = 8h^{-1}\,\mathrm{Mpc}$ is known as σ_8, and is a widely-used cosmological parameter that characterises the amplitude of the present-time large-scale fluctuations in the Universe. Current observations give $\sigma_8 \approx 0.8$ (see Section 1.5).

Linear theory is accurate enough to compute the matter power spectrum on very large scales, and also to compute the CMB power spectrum (Hu and Dodelson, 2002).

It is not, however, sufficient to describe the matter distribution on small scales, e.g., to describe galaxy clusters, which are collapsed, highly nonlinear objects. Roughly, linear theory breaks down when $\sigma(R, a) \sim 1$. However, it is a useful starting point of some interesting analytical approaches to understand galaxy clusters, as we will see in subsequent sections.

1.2.4 The spherical collapse model

The spherical collapse model (Gunn and Gott, 1972) is the simplest model of galaxy cluster formation. Though very limited, it is the basis of insightful models of cluster abundance and clustering (see Sections 1.2.5 and 1.2.6), and it motivates the most commonly used definition of cluster mass.

In its simplest version, the spherical collapse model considers a spherically-symmetric top-hat overdensity in an otherwise smooth, Einstein-de Sitter (EdS) universe, i.e., a spatially flat FLRW universe with $\Omega_\mathrm{m} = 1$, for which $a \propto t^{2/3}$. A possible way to create such an overdensity is to consider a sphere in the smooth background and to radially contract it to produce a homogeneous, overdense sphere, leaving some 'empty space' between it and the expanding background. Due to Birkhoff's theorem (or to Newton's shell theorem), the overdensity and the background evolve independently. Suppose that the overdensity was created at some initial time, t_i, and that is has an initial physical radius, R_i, and an enclosed mass, M, which remains constant, and which, if the radius is known, fixes the density contrast. The physical radius R then evolves with time as the scale factor of a closed $\Omega_\mathrm{m} > 1$ FLRW universe; a parametric solution can be written as (Peacock, 1999)

$$\begin{aligned} R &= A(1 - \cos\theta) \\ t &= B(\theta - \sin\theta), \end{aligned} \tag{1.21}$$

where A and B are determined by the initial conditions, subject to the constraint $A^3 = GMB^2$, which arises from the equation of motion. Expanding R and t up to order θ^2 shows that, initially, $R \propto t^{2/3}$, i.e., the overdensity initially expands like the background. However, it decelerates more quickly, and eventually it reaches a maximum radius at $\theta = \pi$, $R_\mathrm{ta} = 2A$, which corresponds to a turn-around time $t_\mathrm{ta} = \pi B$. At this point, the forming halo has completely detached from the Hubble flow. It then starts to contract, formally collapsing to a singularity at $\theta = 2\pi$, which corresponds to a time $t_\mathrm{vir} = 2t_\mathrm{ta}$. This, of course, never happens in practice, since real

collapse is not perfectly spherically symmetric. The idea is that any deviations from spherical symmetry turn part of the gravitational potential energy of the overdensity into kinetic energy associated with random motions of the particles in it, and the collapsing overdensity reaches virial equilibrium (e.g., Lynden-Bell 1967). The virial theorem allows to calculate the final radius of the collapsed halo, the virial radius, which is $R_{\rm vir} = R_{\rm ta}/2$.

An interesting quantity to consider is the mean overdensity enclosed within the halo as a function of time, which is defined as

$$\Delta(t) = \frac{\rho(t)}{\bar{\rho}_{\rm m}(t)}, \tag{1.22}$$

where $\rho(t)$ is the mean density within the halo, $\rho(t) = M(4\pi R(t)^3/3)^{-1}$, and $\bar{\rho}_{\rm m}(t)$ is the background density, which for an EdS universe can be written as $\bar{\rho}_{\rm m}(t) = (6\pi G t^2)^{-1}$. Expanding this expression up to order θ^5 gives

$$\Delta(t) \simeq 1 + \frac{3}{20}\left(\frac{6t}{B}\right)^{2/3} \equiv 1 + \delta(t). \tag{1.23}$$

At early times, $\delta(t)$ grows as the scale factor; we thus identify it with the linear theory density contrast. At turn-around, $\delta \simeq 1.06$, i.e., linear theory rightly suggests that the evolution is going nonlinear ($\delta \sim 1$), and $\Delta \simeq 5.55$. At virialisation, $\delta \simeq 1.69$, and $\Delta \simeq 178$. That is, a virialised halo with a mean overdensity of $\Delta_{\rm vir} \simeq 178$ forms wherever linear theory predicts a density contrast of $\delta_{\rm c} \simeq 1.689$, known as the critical density contrast. Note that this is the overdensity at the time of collapse. After collapse, the overdensity will increase with time, since the background density decreases with the expansion, whereas the halo mass and virial radius remain constant.

These turn-around and virial overdensities are, remarkably, independent of the mass of the halo and of redshift, although formation times do depend on the initial conditions. However, in a general background cosmology they take different values and are, in general, cosmology and redshift-dependent (see Bryan and Norman 1998 for approximating formulae). In general, $\delta_{\rm c}$ has a weaker dependence on redshift and cosmology than $\Delta_{\rm vir}$ does (Percival, 2005). For $\Omega_{\rm m} = 0.27$ and $\Omega_\Lambda = 0.73$ at $z = 0$, $\delta_{\rm c} \simeq 1.675$ and $\Delta_{\rm vir} \simeq 358$ (Kravtsov and Borgani, 2012), where we note that $\Delta_{\rm vir}$ is with respect to the mean matter density.

The spherical collapse model is, of course, limited. Initial overdensities are not top-hat and spherically symmetric, but smooth and asymmetric (Bardeen et al., 1986). Furthermore, collapse and virialisation do not happen instantaneously. Extensions that consider, e.g., ellipsoidal collapse have been proposed (Icke, 1973; Sheth et al., 2001),

but simulations are needed if the full richness of halo formation is to be captured. Simulations show that collapse is a complex process extended in time, with mergers and constant accretion of matter into halos (Diemand et al., 2007). Furthermore, halos are not isolated objects, but form part of the three-dimensional cosmic web, being connected among them by filaments, along which accretion preferentially takes place (Bond et al., 1996). Moreover, they are not smooth, but present substructure on all scales (Kravtsov and Borgani, 2012).

1.2.5 The halo mass function

As noted in Section 1.1.4, a key ingredient of cluster counts analyses is to compute the halo abundance as a function of mass and redshift in a given cosmology. This abundance can be written as

$$\frac{d^4N}{dMdzd^2\Omega} = \frac{d^4N}{dMd^3V}\frac{d^3V}{dzd^2\Omega}, \tag{1.24}$$

where V denotes comoving volume, and where $\frac{d^4N}{dMd^3V}$ is known as the halo mass function, which depends on mass, redshift, and cosmology; $\frac{d^3V}{dzd^2\Omega}$ is the comoving volume element. For a flat FLRW universe, the volume element can be written as

$$\frac{d^3V}{dzd^2\Omega} = \frac{c\chi(z)^2}{H(z)} = \left(\frac{c}{H_0}\right)^3 \frac{1}{E(z)}\left[\int_0^z \frac{1}{E(z')}dz'\right]^2. \tag{1.25}$$

Therefore, it depends on H_0, and on the density parameters through the normalised expansion rate, $E(z)$.

The spherical collapse model motivated the first attempt to calculate the halo mass function, the Press-Schechter mass function (Press and Schechter, 1974), which is remarkably accurate (about a factor of two) given its simplicity. In practice, halo mass functions calibrated with simulations are used if percent-level accuracy is required.

The Press-Schechter halo mass function

The spherical collapse model suggests that any region where $\delta > \delta_c$, where δ is the linearly extrapolated matter density contrast, has collapsed into a bound halo. The key assumption of the Press-Schechter model is that, at a given time, the fraction of total mass that is in collapsed halos with mass larger than M is equal to the probability that a given point has $\delta_M > \delta_c$, where δ_M is the linear density field filtered on a mass scale M (or, equivalently, its corresponding length scale R, see Eq. 1.19). This is equivalent to assuming that the only halos that exist at a given time are those that have just

collapsed: the regions that correspond to a given mass M are those with $\delta_M = \delta_c$, and regions with $\delta_M > \delta_c$ can be thought of as belonging to halos with higher masses. Assuming Gaussian initial conditions, since linear evolution preserves Gaussianity, this probability can be written as

$$F(M) = \frac{1}{\sqrt{2\pi}\sigma(M)} \int_{\delta_c}^{\infty} e^{-\frac{\delta_M^2}{2\sigma(M)^2}} d\delta_M = \frac{1}{2}\mathrm{erfc}\left(\frac{\nu}{\sqrt{2}}\right), \tag{1.26}$$

where $\sigma(M)$ is the standard deviation of the filtered field, given by Eq. (1.20) (the scale factor dependence has been left implicit), and where ν, the peak height, is defined as

$$\nu \equiv \frac{\delta_c}{\sigma(M)}. \tag{1.27}$$

Press and Schechter further assumed that all matter is contained within collapsed objects, in which case one should expect $F(0) = 1$. However, $\nu \to 0$ as $M \to 0$, and therefore $F(0) = 1/2$. A possible solution, and indeed what was done in Press and Schechter (1974), is to simply introduce an *ad hoc* factor of 2.

A more principled justification of this factor of 2 was proposed in Bond et al. (1991). The idea is that a given point may have $\delta_{M_1} < \delta_c$ on a given scale M_1, but may have $\delta_{M_2} > \delta_c$ on a larger scale $M_2 > M_1$. In this scenario, the point does belong to a collapsed halo with $M > M_1$, since $M_2 > M_1$, but the Press-Schechter *ansatz* is missing it. What has to be considered instead is the largest scale on which the threshold, δ_c, is crossed. This can be done by considering random walks in $(\sigma(M), \delta_M)$ space, starting at an arbitrarily large mass, and, thus, starting with $\sigma(M)$ arbitrarily close to 0. In this approach, $\sigma(M)$, which decreases monotonically with M, acts as the 'time' variable. The key assumption now is that the fraction of total mass which is in collapsed halos with mass larger than M is the fraction of random walks that have crossed the threshold, δ_c, for the first time at any point before $\sigma(M)$, i.e., at any larger mass. The only analytically-tractable case is that in which the filter used is a top-hat filter in Fourier space (Bond et al., 1991). With such a choice of filter, as the resolution is increased, i.e., as mass is decreased, new, independent modes are simply added to the old ones: the random walk is Markovian. At a given mass M, the fraction of trajectories that is currently above the threshold is given by Eq. (1.26) (indeed, this corresponds to the mass fraction according to the Press-Schechter original assumption). Those trajectories have crossed the threshold at some earlier time (i.e., larger mass), and, at the moment of threshold crossing, for every random walk that remains above it, there is a mirror random walk that crosses it back with

equal probability (Chandrasekhar, 1943). Therefore, the fraction of random walks that have ever crossed the threshold at any mass larger than M is just twice what is given in Eq. (1.26).

Whatever the argument to obtain $F(M)$, the fraction of mass in halos between M and $M + dM$ is then $\left|\frac{dF}{dM}\right| dM$, and thus

$$\frac{d^4N}{dMd^3V} = \frac{\bar{\rho}_{\rm m,0}}{M}\left|\frac{dF}{dM}\right|, \tag{1.28}$$

which, for the Press-Schechter *ansatz* (factor of 2 included) can be written as

$$\frac{d^4N}{dMd^3V} = \sqrt{\frac{2}{\pi}}\frac{\Omega_{\rm m}\rho_{\rm c,0}}{M^2}\frac{\delta_{\rm c}}{\sigma(M)}\left|\frac{d\ln\sigma(M)}{d\ln M}\right| e^{-\frac{\nu^2}{2}} = \frac{\Omega_{\rm m}\rho_{\rm c,0}}{M^2}\left|\frac{d\ln\sigma(M)}{d\ln M}\right| f(\sigma(M)), \tag{1.29}$$

where $f(\sigma)$ describes the shape of the mass function.

This mass function deviates by about a factor of 2 from simulations, over-predicting low mass halos and under-predicting high mass halos (e.g., Jenkins et al. 2001). Other mass functions have been proposed along similar lines, based, e.g., on ellipsoidal collapse (e.g., Sheth and Tormen 2002). These have greater, but still insufficient, accuracy to match the current percent-level halo mass functions directly calibrated with simulations.

Halo mass function from simulations

A key factor in determining the halo mass function from simulations, and in galaxy cluster science in general, is the definition of halo (or cluster) mass. Halos in simulations, and, indeed, in reality, are extended, triaxial objects whose boundaries are not well-defined, and therefore there is no unique way to define their masses. The most commonly used definition is the spherical overdensity mass, which is defined as follows. First, a halo centre must be defined. In a simulation, this can be done by finding, e.g., the most bound particle (Kravtsov and Borgani, 2012). The spherical overdensity mass is then defined as the total mass within a sphere centred on the halo centre within which the mean density is $\Delta_{\rm ref}$ times a reference density, $\rho_{\rm ref}$. This reference density is typically either the critical or the mean matter density of the Universe at the halo's redshift. This defines a halo mass, M_Δ, and a halo radius, R_Δ, with

$$M_\Delta = \frac{4}{3}\pi\Delta_{\rm ref}\rho_{\rm ref}(z)R_\Delta^3, \tag{1.30}$$

Several values of $\Delta_{\rm ref}$ (of both $\Delta_{\rm mean}$ and $\Delta_{\rm critical}$) are used in practice. The choice is usually motivated by observations. For instance, X-ray observations are typically restricted to the inner regions of halos, so large values of $\Delta_{\rm ref}$, up to $\Delta_{\rm critical} \sim 2500$, are common, whereas lensing observations, which extend into the halo outskirts, typically measure $\Delta_{\rm critical} \sim 200$–$500$ (Kravtsov and Borgani, 2012). Note that $\Delta_{\rm mean} = \Delta_{\rm critical}/\Omega_{\rm m}(z)$, and that direct conversion between masses at different overdensities is not possible, unless a density profile is assumed (see, e.g., Hu and Kravtsov 2014).

The spherical overdensity mass definition is generally preferred over other existing definitions, e.g., the friends-of-friends mass, which presents some problems such as assigning a single halo to two observationally distinct clusters. It also correlates better with observed quantities that are integrated over an aperture (Allen et al., 2011). A downside is that sometimes two spheres can overlap, leading to double-counting of mass, although this does not happen frequently (Tinker et al., 2008).

Currently, halo mass functions from simulations are determined to percent accuracy. In particular, Tinker et al. (2008) used 22 large N-body simulations to calibrate the halo mass function with 5–10 % precision from $z = 0$ to $z = 2$ for a ΛCDM cosmology for a number of overdensities between $\Delta_{\rm mean} = 200$ and $\Delta_{\rm mean} = 3200$. They parametrise the halo mass function with a universal function of $\sigma(M)$ (i.e., cosmology enters the halo mass function only through $\sigma(M)$),

$$f(\sigma) = A\left[\left(\frac{b}{\sigma}\right)^a + 1\right] e^{-\frac{c}{\sigma^2}}, \tag{1.31}$$

where A, a, b, and c are fitting parameters that are only mildly redshift-dependent. Compare this fitting function with the Press-Schechter analogous function in Eq. (1.29); in both cases there is an exponential suppression at high masses (i.e., small values of σ).

Recently, Bocquet et al. (2020) followed an alternative approach to determining the halo mass function by 'emulating' it from N-body simulations using Gaussian processes, achieving percent-level accuracy on cluster scales.

Calibrating halo mass functions with N-body simulations can be done, in principle, to arbitrary precision. The effect of baryons, due to, e.g., supernovae and AGN feedback, which can lead to some redistribution of mass in halos, is, however, more difficult to quantify, with predictions not being completely robust yet. In general, baryons are thought to affect the halo mass function to a few percent (Kravtsov and Borgani, 2012), and whether this has a significant impact on cosmological analyses depends on the particular sample. For instance, using N-body dark-matter-only simulations and

simulations with baryons, Bocquet et al. (2016) concluded that baryonic corrections are not significant for *Planck* and SPT-like SZ cluster samples, but should be taken into account for an eROSITA-like X-ray sample.

Dependence on cosmology and on mass calibration

Assuming a flat ΛCDM model, cluster counts are very sensitive to $\Omega_{\rm m}$ and σ_8. This dependence arises from the halo mass function through the linear growth factor, to which there is exponential sensitivity (see Eqs. 1.29 and 1.31), and from the volume element through the expansion rate (see Eq. 1.24). If the cluster sample considered is local (i.e., at low redshift), cluster counts offer a direct measurement of σ_8, as this parameter acts as a normalisation of the linear growth function at $z = 0$. As redshift increases, the linear growth function, and thus cluster counts, becomes increasingly sensitive to $\Omega_{\rm m}$, which helps to break the degeneracy with σ_8 (see, e.g., Voit 2005). A number of competitive constraints on $\Omega_{\rm m}$ and σ_8 have been derived from cluster counts; see Section 1.5. We also present constraints on these two parameters in Chapter 3 obtained from *Planck* SZ-detected clusters with CMB lensing mass calibration.

There is also sensitivity to spatial curvature (e.g., Planck 2015 results XXIV 2016), and to possible deviations from the cosmological constant through their impact on the growth function and the expansion rate. For example, the equation of state of dark energy, w, can be constrained along with $\Omega_{\rm m}$ (or, equivalently, $\Omega_{\rm DE}$) assuming a flat wCDM cosmology (e.g., Mantz et al. 2015; Planck 2015 results XXIV 2016).

Cluster counts are not very sensitive to other cosmological parameters, such as H_0, n_s, and, especially, $\Omega_{\rm b}$. Thus, additional information must be used in order to constrain these parameters if they are allowed to vary in counts analyses (see, e.g., Chapter 3).

Finally, as with σ_8, in the halo mass function there is exponential sensitivity to cluster mass. Indeed, $\sigma(M)$ is present in the exponential term of the halo mass function, and $\sigma(M)$ is roughly a power law of halo mass (since the primordial power spectrum can be approximated locally around cluster scales by a power law; see, e.g., Kravtsov and Borgani 2012). This exponential sensitivity means that any biases in the determination of halo masses propagate exponentially into the predicted halo abundance, especially at high masses, where the exponential term dominates. Making sure that cluster masses are unbiased is therefore crucial for the success of cluster counts analyses.

1.2.6 Clustering of clusters

Galaxy clusters are not randomly distributed across the Universe, but they tend to cluster with each other. They are in fact strongly clustered, more than galaxies themselves (Bahcall and Soneira, 1983). This clustering can be parametrised with the linear bias, b (Kaiser, 1984), which relates the local density contrast of halos, $\delta_{\rm h}(\mathbf{x}) \equiv (n_{\rm h}(\mathbf{x}) - \bar{n}_{\rm h})/\bar{n}_{\rm h}$, where $n_{\rm h}(\mathbf{x})$ and $\bar{n}_{\rm h}$ denote, respectively, the local and mean halo number density, to the local matter density contrast, $\delta(\mathbf{x})$,

$$\delta_{\rm h}(\mathbf{x}) = b\delta(\mathbf{x}). \tag{1.32}$$

The linear bias, b, depends on halo mass, increasing with increasing mass: the more massive the halo, the more biased with respect to the matter distribution it is (e.g., Tinker et al. 2010). However, it is approximately independent of scale for Gaussian initial conditions (Scherrer and Weinberg, 1998). It does depend on scale if initial conditions are non-Gaussian, which makes it possible to use measurements of the linear bias in order to constrain some forms of primordial non-Gaussianity (Dalal et al., 2008).

The linear bias is closely related to the halo mass function, and can in fact be derived from it within the peak-background split formalism (Kaiser, 1984). This approach yields predictions of the linear bias that are accurate to about 20 % (Tinker et al., 2010). More precise determinations are obtained by direct calibration with simulations. For example, Tinker et al. (2010) calibrated the linear bias as a universal function of ν, $b = b(\nu)$, to better than 5 % accuracy, with fitting parameters only dependent on the mass overdensity $\Delta_{\rm ref}$ chosen.

The effect of clustering of galaxy clusters on analyses of their counts depends on the sample considered. For a *Planck*-like sample, Poisson noise dominates and the covariance due to clustering can be neglected (Planck 2015 results XXIV, 2016). For other samples, however, it has to be taken into account (see, e.g., Hu and Kravtsov 2003).

1.2.7 Halo internal structure

As already noted, real halos are not top-hat overdensities, as assumed in the spherical collapse model. The spherically-averaged halo density profile is commonly approximated with the Navarro-Frenk-White (NFW) profile (Navarro et al., 1995, 1997),

$$\rho_{\rm NFW}(x) = \frac{4\rho_s}{x(1+x)^2}, \tag{1.33}$$

where $x \equiv r/r_s$, and r_s is a characteristic radius, and where ρ_s is the density at the characteristic radius. This profile is a good approximation of the spherically-averaged density from N-body simulations; remarkably, it is a universal function of x. Its density diverges at $x = 0$, but the enclosed mass within some value of x is finite; however, the enclosed mass diverges logarithmically with x. Several modifications to the NFW profile exist that guarantee a finite total enclosed mass at the expense of additional parameters: for example, the profile can be simply truncated at some radius (e.g., Melin and Bartlett 2015), or smoothly driven to zero in a faster way (e.g., Baltz et al. 2009).

The characteristic radius, r_s, is related to the spherical overdensity radius, R_Δ, where Δ is the overdensity of choice, through the concentration parameter, c_Δ, $R_\Delta = c_\Delta r_s$. Assuming the NFW profile, a halo is completely specified by its spherical overdensity mass, M_Δ (or, equivalently, R_Δ), its redshift, and the concentration parameter, c_Δ. The concentration parameter depends, in general, on mass, redshift, and cosmology; at fixed redshift, it decreases with mass (e.g., Navarro et al. 1997; Diemer and Kravtsov 2015). Recently, Diemer and Kravtsov (2015) calibrated c_Δ with N-body simulations as a universal function of ν, $c_\Delta = c_\Delta(\nu)$, up to 5 % accuracy for ΛCDM.

Another popular profile is the Einasto profile (Einasto, 1965), which has recently been found to provide a better description of the mean halo profile at the expense of an additional parameter, α (e.g., Gao et al. 2012).

Halos are, of course, not spherically symmetric, but triaxial objects living in the cosmic web. Relaxed halos are typically less triaxial. Thus, triaxiality generally increases with both redshift and mass, since massive halos form at later times (e.g., Kasun and Evrard 2005).

1.2.8 From halos to galaxy clusters: baryons

Dark matter is the main component of galaxy clusters, and, as we have seen, cluster formation, abundance, spatial distribution, and internal structure can be understood quite well as just the result of gravity acting on dark-matter-only objects. However, as already noted in Section 1.1.2, galaxy clusters also contain baryonic matter, both in the form of galaxies and of diffuse gas in the ICM.

The main driving force governing a cluster's baryonic properties is, unsurprisingly, gravity. This makes it possible to predict, to some extent, the scalings of some of these properties with cluster mass and redshift with the framework of the self-similar model (Kaiser 1986; see also Kravtsov and Borgani 2012). However, the full understanding of baryons in clusters, taking into account all the wealth of baryonic physics involved,

is a much more difficult task. Except in the vicinity of black holes, this physics is well-understood classical physics. Nevertheless, the processes involved are highly nonlinear, and there is a huge dynamic range in the relevant length and time scales, from supernovae, which take place on a stellar scale but which can affect a whole cluster and its environment, to the cosmic-scale ICM (Allen et al., 2011). This makes both the theoretical understanding of baryons in clusters and their incorporation to simulations very challenging problems.

As the gas falls into the forming cluster gravitational well, its gravitational potential energy is turned into kinetic energy, which in turn thermalises via adiabatic compression and shocks. These shocks are due to supersonic flows associated mostly to accretion of matter, in which case they are typically present at the cluster boundary, where they can reach Mach numbers of 30 and above (e.g., Vazza et al. 2009). There can also be smaller shocks associated to mergers, and to jets from AGN and stellar wind from star-forming galaxies (Allen et al., 2011). After some time, equilibrium can be attained, with the ICM reaching its virial temperature. In general, thermalisation is quite efficient, with most of a cluster's ICM pressure being of thermal origin (Kravtsov and Borgani, 2012). There can be, however, significant sources of non-thermal pressure, as we will discuss in Section 1.3.1.

The main source of non-gravitational energy is AGN feedback, which heats up the ICM and redistributes the gas at small radii (e.g., Le Brun et al. 2014). It is particularly important at low masses, $M < 10^{14} M_{\odot}$ (Pratt et al., 2019). Another, less important, non-gravitational energy source is feedback from supernovae, which also heats up the ICM (Le Brun et al., 2014). Acting in the opposite direction is ICM cooling, a process that is necessary for star formation. The main source of cooling is radiative cooling, which is a relatively inefficient process, with timescales typically longer than the Hubble time (Allen et al., 2011). This is evidenced by the small fraction of cluster mass that is in the form of galaxies. An exception are the central, 0.1 Mpc regions of some relaxed clusters, which present dense, very X-ray bright cool cores (Allen et al., 2011).

Radiative cooling and feedback from AGN and from other compact sources are tightly coupled, forming a 'self-regulating' cycle in which cooling leads to star and AGN formation, which in turn suppress further cooling (Kravtsov and Borgani, 2012). Trying to understand fully these complex processes is currently at the forefront of galaxy cluster astrophysics.

1.3 Mass estimates and mass proxies

As noted in Section 1.1.4, cluster mass determination is currently the main source of systematic uncertainty in cluster counts analyses, limiting their statistical power (Pratt et al., 2019). Due to this, there is a huge ongoing effort to understand and control systematics in mass estimation, including the important task of carefully comparing cluster masses obtained by different groups using different methodology (see, e.g., Sereno and Ettori 2015). In this section we review some of the different ways in which cluster masses are estimated, and also some of the various mass proxies that are used in counts analyses, highlighting possible sources of systematic error. We do not consider CMB lensing cluster mass estimation, which we discuss in Section 1.4.

1.3.1 Equilibrium?

When a galaxy cluster forms, after a few sound crossing times (typically around 10^9 years each), it is expected to reach equilibrium (Pratt et al., 2019). For the collisional gas that comprises the ICM, this means hydrostatic equilibrium, i.e., that the pressure gradient balances the gravitational pull. Assuming spherical symmetry, this condition can be written as (Kravtsov and Borgani, 2012)

$$\frac{dp(r)}{dr} = -G\frac{\rho_{\rm g}(r)M(<r)}{r^2}, \tag{1.34}$$

where p is the gas pressure, $M(<r)$ is the mass within a radius r, and $\rho_{\rm g}$ is the gas density. If the ICM gas has completely thermalised, the only significant source of pressure is gas thermal pressure. Under this assumption, and also assuming that the gas can be described by the ideal gas equation of state, $p = \rho_{\rm g}k_{\rm B}T/(\mu m_{\rm p})$, where T is the gas temperature, μ is the mean molecular weight, and $m_{\rm p}$ is the proton mass, Eq. (1.34) can be rewritten as

$$M_{\rm HE}(<r) = -\frac{rk_{\rm B}T(r)}{G\mu m_{\rm p}}\left[\frac{d\ln\rho_{\rm g}(r)}{d\ln r} + \frac{d\ln T(r)}{d\ln r}\right]. \tag{1.35}$$

That is, the cluster mass within a given radius r can be obtained if, at such radius, the gas temperature and the derivatives of the gas temperature and of the gas density are known. This can be done, e.g., with detailed X-ray observations. We denote such estimated mass with $M_{\rm HE}(<r)$, highlighting that hydrostatic equilibrium is a key assumption.

Thermal pressure is known to be the main source of pressure in most clusters, especially in relaxed ones that have not undergone recent mergers (Pratt et al., 2019).

Nevertheless, there can be other substantial contributions to the total pressure with a non-thermal origin, which can significantly bias mass estimates based on Eq. (1.35). The assumption of hydrostatic equilibrium may not be valid either, especially at large radii and for clusters that have recently experienced major mergers. Currently there is a large ongoing effort towards quantifying these additional sources of pressure and possible departures from equilibrium, and towards determining their impact on mass estimation (e.g., Rasia et al. 2006; Vazza et al. 2009; Lau et al. 2009; Battaglia et al. 2012; Nelson et al. 2014a,b; Biffi et al. 2016; Shi et al. 2016). This is mostly done with simulations, in which it is consistently found that random, turbulent motions are the main source of non-thermal pressure, independently from the cluster's dynamical state. Turbulent pressure, which adds to the thermal pressure, $p_{\rm th}$, as $p_{\rm total} = p_{\rm th} + \frac{1}{3}\rho_{\rm g}\sigma^2$, where σ is the (isotropic) random velocity dispersion, can account for about 10–30 % of the total pressure at R_{500} (Nelson et al., 2014a; Shi et al., 2016). Other sources of bias include bulk motions and acceleration, which are important in merging clusters and at large radii (Lau et al., 2009; Nelson et al., 2014b), AGN feedback, which mostly affects the central regions (Allen et al., 2011), and inhomogeneities in the gas distribution, which can introduce a $\sim 5\,\%$ bias if the hydrostatic equilibrium equation is not modified to account for them (Roncarelli et al., 2013).

The total bias in mass estimation can be summarised with the so-called hydrostatic mass bias parameter, which relates the true cluster mass to its mass estimated through Eq. (1.35) or through mass proxies calibrated with hydrostatic masses. This hydrostatic mass bias can also account for other systematics in the measurements, such as the instrument absolute X-ray calibration, thus becoming an effective mass bias. This is the case of the *Planck* clusters mass bias used in Planck 2015 results XXIV (2016), and which we calibrate in the cosmological analysis presented in Chapter 3.

An analogous equation to Eq. (1.35) can be written for the collisionless cluster galaxies and dark matter: the Jeans equation. Assuming spherical symmetry and absence of rotation (most clusters do not rotate significantly, e.g., Hwang and Lee 2007), it can be written as (Kravtsov and Borgani, 2012)

$$M_{\rm J}(<r) = -\frac{r\sigma_r^2}{G}\left[\frac{d\ln\nu(r)}{d\ln r} + \frac{d\ln\sigma_{\rm r}(r)^2}{d\ln r} + 2\beta(r)\right], \tag{1.36}$$

where ν is the number density (e.g., of galaxies), $\sigma_{\rm r}$ is the radial component of the velocity dispersion, and $\beta \equiv 1 - \frac{\sigma_{\rm t}^2}{2\sigma_{\rm r}^2}$, where $\sigma_{\rm t}$ is the tangential component of the velocity dispersion, is the anisotropy parameter, which vanishes if the velocity distribution is isotropic. Thus, analogously to the hydrostatic equilibrium equation, the cluster

mass within some radius r can be obtained if, at such radius, the particle (e.g., galaxy) velocity dispersion and the derivatives of the velocity dispersion and of the number density are known.

Cluster masses estimated with Eq. (1.36) can also be biased if dynamical equilibrium is violated. This bias can be as large as a factor of 2 for major mergers (Takizawa et al., 2010), although departures from equilibrium can be identified from the velocity distribution of the cluster galaxies (Pratt et al., 2019).

1.3.2 X-ray observations

Mass estimation from X-ray observations typically relies on the hydrostatic equilibrium equation (Eq. 1.35). The gas density profile is usually derived from X-ray surface brightness measurements azimuthally averaged over thin, concentric annuli, and the gas temperature profile can be obtained from spectra measurements, also averaged over thin annuli. Several approaches exist to obtain the cluster mass profile from these measurements, which range from fully parametric to completely non-parametric methods (see, e.g., Ettori et al. 2013). Alternatively, the pressure profile can be obtained from azimuthally averaged SZ measurements, in which case Eq. (1.34) is used and temperature measurements are not needed (e.g., Ameglio et al. 2009).

The most significant source of bias in these mass measurements is thought to be the assumption of hydrostatic equilibrium. Indeed, simulations suggest that this assumption typically biases low mass estimates by about 20 % (Pratt et al., 2019). Other sources of bias include triaxiality (a few %; Buote and Humphrey 2012), temperature inhomogeneities within the annuli (around 10–15 %; Rasia et al. 2014), and absolute X-ray instrumental calibration (up to 15–20 %; Schellenberger et al. 2015). As we will see in Section 1.5, some recent analyses have found a large total mean X-ray mass bias, of up to about 40 % (e.g., Planck 2015 results XXIV 2016). The origin of such a large value has, so far, remained unclear.

Accurately determining the scaling relations of X-ray mass proxies is equally important to obtaining unbiased mass estimates, especially in the context of cosmological analyses. Recently there has been significant convergence in the calibration of scaling relations of X-ray mass proxies (Pratt et al., 2019). Frequently-used proxies include the total X-ray luminosity, $L_{\rm X}$, the gas mass, $M_{\rm g}$, the average gas temperature, $T_{\rm X}$, and $Y_{\rm X}$, which is defined as $Y_{\rm X} = M_{\rm g} T_{\rm X}$, and is therefore a measure of the ICM thermal energy (Kravtsov et al., 2006). The main source of scatter in all these proxies is intrinsic. $L_{\rm X}$ has large scatter (around 40 %), unless the core is excised (less than about 10 %; e.g.,

Mantz et al. 2018), whereas $M_{\rm g}$, $T_{\rm X}$, and $Y_{\rm X}$ have smaller scatter (about 10–15 %; e.g., Kravtsov et al. 2006; Arnaud et al. 2007; Allen et al. 2008).

1.3.3 Optical observations: kinematics

A widely used way to measure cluster masses with optical observations is based on Eq. (1.36), using the cluster galaxies as particles, and with their number density and velocity dispersion inferred from their observed number density and line-of-sight velocities (e.g., Carlberg et al. 1997). This approach is insensitive to some sources of non-thermal pressure that bias X-ray mass estimates, such as turbulent motions, but it assumes dynamical equilibrium, with galaxies relaxing on a longer timescale than the ICM (White et al., 2010). As already noted, this assumption can introduce a significant bias a large as a factor of 2 (Takizawa et al., 2010). Another source of bias is triaxiality (around 30 %), which can be reduced by stacking (Pratt et al., 2019). This method is also limited by the finite number of galaxies in a cluster, and by the fact that it is more difficult to find cluster centres with optical observations than it is in X-ray analyses (Allen et al., 2011). An alternative method that does not rely on the assumption of equilibrium exists, the Caustic method (Diaferio and Geller, 1997), which is applicable to regions beyond the virial radius (see, e.g., Biviano and Girardi 2003).

A commonly used mass proxy is the velocity dispersion. Although it has small intrinsic scatter (≤ 5 %), observational uncertainty induces a total scatter of about 40 % (White et al., 2010). Another common proxy is the richness, the number of galaxies within a given aperture, which has a roughly similar scatter of about 40 % (Rozo et al., 2010). Currently, these large scatter values, as well as systematics, strongly limit the statistical power of optical cluster counts analyses, which typically use a larger number of clusters than their X-ray and SZ counterparts (Pratt et al., 2019).

1.3.4 Optical observations: weak lensing

Over the past decade, galaxy weak lensing has become the most robust way to obtain galaxy cluster mass estimates (Pratt et al., 2019). Also, recently, shear-selected cluster catalogues have been constructed (e.g., Miyazaki et al. 2018). The key property of lensing is that it is completely independent of the physical properties and of the dynamical state of clusters: it is only sensitive to their mass distribution.

In weak lensing, the small, but coherent distortions of the shapes of background galaxies are used in order to estimate the lensing shear, γ, which is given by the

lens (galaxy cluster, in this case) mass distribution (see, e.g., Bartelmann 2010). In particular, the observed galaxy complex ellipticity can be written as $\epsilon = \epsilon_{\rm g} + g$, where $\epsilon_{\rm g}$ is the galaxy intrinsic ellipticity (known as 'shape noise'), and g is known as the reduced shear, $g = \gamma/(1-\kappa)$, where κ is the lensing convergence. In the weak limit, $g \simeq \gamma$, and if galaxies are assumed to be oriented randomly, their observed ellipticities, ϵ, become unbiased estimators of the shear, γ. Thus, one can ensemble average the ellipticities of several galaxies to obtain local estimates of the shear. Then, a shear model, which must account for all the relevant covariance (e.g., Gruen et al. 2015), is typically fit to the azimuthally averaged estimated shear to finally produce a mass estimate.

Despite being insensitive to the cluster's dynamical state, lensing is not free from limitations and biases. Important sources of statistical limitation are the number of background galaxies, especially at high redshift, and triaxiality, which can produce an intrinsic scatter in the estimated masses ranging from about 5 % to about 30 % (Becker and Kravtsov, 2011) (note that the dark matter distribution, which is responsible for most of the lensing signal, is more triaxial than the gas distribution, which is responsible for the X-ray and SZ signals). However, systematics currently dominate the error budget in weak lensing analyses (McClintock et al., 2019). The dominant systematic is the selection of background galaxies. The most common case is that in which faint cluster galaxies are mistaken for background galaxies, which leads to an underestimation of the lensing effect and which can, in turn, bias the estimated mass low by up to about 40 % (Okabe and Smith, 2016). Other sources of bias include errors in the cluster lensing signal modelling (up to about 10 %; Pratt et al. 2019), errors in galaxy shape measurements (a few %; Hamana et al. 2013), and errors in galaxy redshift measurements, which are typically photometric (up to a few %; Sánchez et al. 2014).

In recent years, several projects have measured the masses of a total of ~ 100 galaxy clusters through their weak lensing signatures with various ground-based telescopes. These projects include LoCuSS (Okabe et al., 2010), CCCP (Hoekstra et al., 2012), CLASH (Umetsu et al., 2014), and WtG (von der Linden et al., 2014). In general, the mass estimates from these different projects are in very good agreement, save perhaps those from WtG, which are somewhat (about 10–15 %) larger (Pratt et al., 2019). Some of these mass estimates have been used to set the absolute cluster mass scale in counts analyses. This is the case, for instance, of the *Planck* SZ counts analysis, which employed weak-lensing masses from CCCP and WtG for two of its three baseline mass scale calibrations (Planck 2015 results XXIV 2016; see also Chapter 3).

1.3.5 The Sunyaev-Zel'dovich effect

As noted in Section 1.1.2, the thermal SZ (tSZ) effect due to galaxy clusters is a spectral distortion of the CMB blackbody spectrum due to the interaction of the CMB photons with the electrons of the ionised ICM via inverse Compton scattering, which leads to an intensity decrement at frequencies below 217 GHz, and to an intensity increment at frequencies above this cross-over frequency (Sunyaev and Zeldovich 1972; see, e.g., Rephaeli 1995a, Carlstrom et al. 2002, and Mroczkowski et al. 2019 for reviews). The tSZ effect is not only produced by galaxy clusters. For example, the tSZ signal from the gas in cosmic web filaments has recently been detected at low significance (de Graaff et al., 2019). However, clusters, as the hottest, densest large-scale objects in the Universe, are the brightest tSZ source.

In the simplest and most commonly-assumed scenario, the ICM electrons are taken to be non-relativistic and in thermal equilibrium, i.e., distributed according to the Maxwell-Boltzmann distribution. In this case, the distortion is independent of the electron temperature, and can be written in terms of the effective CMB temperature as

$$\frac{\Delta T_{\rm CMB}}{T_{\rm CMB}} \simeq y_{\rm tSZ} \left(x \frac{e^x + 1}{e^x - 1} - 4 \right), \tag{1.37}$$

where $y_{\rm tSZ}$ is the Compton-y parameter, which controls the overall amplitude of the effect, and where $x \equiv h\nu/(k_{\rm B}T_{\rm CMB})$, where ν is the frequency, and $T_{\rm CMB}$ is the present-time CMB temperature.

The Compton-y parameter is given by

$$y_{\rm tSZ} = \int \frac{k_{\rm B}T_{\rm e}}{m_{\rm e}c^2} d\tau_{\rm e} = \int \frac{k_{\rm B}T_{\rm e}}{m_{\rm e}c^2} n_{\rm e}\sigma_{\rm T} dl = \frac{\sigma_{\rm T}}{m_{\rm e}c^2} \int p_{\rm e} dl. \tag{1.38}$$

Here, $T_{\rm e}$ is the electron temperature; $m_{\rm e}$ is the electron mass; $d\tau_{\rm e}$ is the optical depth element along the line of sight, which is $d\tau_{\rm e} = n_{\rm e}\sigma_{\rm T}dl$, where $n_{\rm e}$, $\sigma_{\rm T}$, and dl are, respectively, the electron number density, the Thomson scattering cross-section, and the proper distance element along the line of sight; and $p_{\rm e}$ is the electron pressure, $p_{\rm e} = n_{\rm e}k_{\rm B}T_{\rm e}$. For a typical cluster, $y_{\rm tSZ} \sim 10^{-4}$ (Mroczkowski et al., 2019), which means that the tSZ effect is comparable in amplitude to the CMB primary anisotropies.

As is apparent in Eq. (1.38), the tSZ effect is a measure of the integrated pressure along the line of sight. This makes it an interesting probe of the ICM. For instance, as noted in Section 1.3.2, detailed tSZ measurements can be combined with X-ray-determined gas density profiles in order to obtain hydrostatic cluster mass estimates via Eq. (1.34), without the need to obtain spectroscopic temperatures (e.g., Ameglio et al. 2009). Some other virtues of the tSZ effect include that $y_{\rm tSZ}$ does not decrease

with redshift, unlike luminosity-based observables, and its linear dependence on the gas (electron) density, which implies that it can be observed further out from the cluster centre than the X-ray surface brightness, which scales as the gas density squared (Mroczkowski et al., 2019).

A widely-used cluster mass proxy, $Y_{\rm SZ}$, can be constructed by integrating $y_{\rm tSZ}$ over an aperture or over a volume. For this, a pressure model is usually assumed (and is indeed necessary if the integral is over a volume). Often, the ICM gas is assumed to follow a generalised NFW profile (e.g., Mroczkowski et al. 2009; first proposed in Nagai et al. 2007b). In this case, the pressure profile can be written as

$$p(x) = \frac{p_0}{(c_{500}x)^\gamma \left[1 + (c_{500}x)^\alpha\right]^{\frac{\beta-\gamma}{\alpha}}}, \tag{1.39}$$

where, as in our discussion of the NFW profile, $x \equiv r/r_s$, where $r_s = R_{500}/c_{500}$, and where 500 refers to the overdensity with respect to the critical density at the cluster's redshift.

The three parameters γ, β, and α, quantify, respectively, the slope at $r \ll r_s$, the slope at $r \gg r_s$, and the quickness of the transition between these two slopes around $r \approx r_s$. These parameters have been measured in various studies (e.g., Arnaud et al. 2010), and the associated profile and mass proxy, $Y_{\rm SZ}$, have been used in several cosmological analyses. For example, the counts analysis in Planck 2015 results XXIV (2016) was based on $Y_{\rm SZ}$ measurements in which this 'universal' pressure profile, as calibrated in Arnaud et al. (2010), was assumed.

Since $y_{\rm tSZ}$ does not decrease significantly with redshift, SZ surveys can detect clusters and measure $Y_{\rm SZ}$ to relatively high redshift, which enhances the constraining power of counts analyses by measuring the growth of structure as a function of time. In addition, $Y_{\rm SZ}$ is easy to measure, is relatively insensitive to the cluster dynamical state, and has low scatter at given cluster mass, about 10–15 % (Mroczkowski et al., 2019). However, it is not free from systematic uncertainties, of which the overall scaling of $Y_{\rm SZ}$ with cluster mass is the main source (e.g., de Haan et al. 2016; Planck 2015 results XXIV 2016; Bocquet et al. 2019; Hilton et al. 2018). Other possible systematic issues include contamination from the emission from radio galaxies within the galaxy cluster, thought to be at the few percent level or less (Sehgal et al., 2010), and relativistic corrections. Indeed, the tSZ distortion in Eq. (1.38) assumes that the ICM electrons are non-relativistic. However, typical clusters have temperatures that imply $v_{\rm e}/c \sim 0.1$–0.2, which suggests that relativistic corrections should be important. These can be computed (see Rephaeli 1995b; Stebbins 1997; Challinor and Lasenby

1998; Itoh et al. 1998), and, in general, introduce a dependence of the tSZ spectral distortion on the electron temperature. Neglecting them has been argued to bias low the constraint on σ_8 from the *Planck* tSZ power spectrum, which did not include them, by about $1\,\sigma$ (Remazeilles et al., 2019). SZ counts analyses from *Planck* and similar experiments could, potentially, also be sensitive to these corrections (Remazeilles et al., 2019).

In addition to the thermal SZ effect, which is due to the thermal motion of the ICM electrons, there is also a SZ signal associated with the peculiar motions of clusters. This is known as the kinetic SZ (kSZ) effect (Sunyaev and Zeldovich, 1980), and can be seen as a Doppler shift contribution to the CMB due to cluster bulk motion. As such, it has the same spectral dependence as the primary CMB (a blackbody), and, as a consequence, cannot be distinguished from the primary anisotropies unless additional information is used (e.g., scale information, or correlations with other probes). The kSZ signal can be written in terms of the effective CMB temperature as

$$\frac{\Delta T_{\rm CMB}}{T_{\rm CMB}} \simeq -\int \frac{\hat{\mathbf{n}}\cdot\mathbf{v}_{\rm p}}{c} d\tau_{\rm e} = -\int \frac{\hat{\mathbf{n}}\cdot\mathbf{v}_{\rm p}}{c} \sigma_{\rm T} n_{\rm e} dl \equiv -y_{\rm kSZ}, \tag{1.40}$$

where $\hat{\mathbf{n}}$ is the unit vector specifying the line of sight, and $\mathbf{v}_{\rm p}$ is the cluster's peculiar velocity. Typical cluster peculiar velocities are a few hundreds of $\mathrm{km\,s^{-1}}$, which means $y_{\rm kSZ} \sim 10^{-5}$ (Mroczkowski et al., 2019), which is about an order of magnitude smaller than the typical tSZ signal (for merging clusters along the line of sight, however, the kSZ signal can dominate over the tSZ signal during core passage; Ruan et al. 2013).

The kSZ effect is an interesting source of information in its own right, from which, e.g., cluster velocities can be extracted (e.g., Haehnelt and Tegmark 1996; Hand et al. 2012). However, it can also be an unwanted foreground, as is the case in CMB lensing cluster mass estimation, where a cluster's kSZ signal cannot be easily separated from the lensed primary CMB anisotropies, since they both have the same spectral signature.

Cluster rotation also produces an SZ signal, which, analogously to the kSZ effect, can be seen as a Doppler shift contribution, in this case due to rotation. This signal has been recently detected at low significance (Baxter et al., 2019). It has a dipole-like structure, as does cluster CMB lensing (see Section 1.4.4), so could potentially be an issue in cluster lensing analyses. However, the direction of the rotation SZ dipole is not aligned with the background gradient of the CMB temperature, as is the CMB lensing dipole, but with the rotation axis, so rotation should not introduce a bias in statistical (i.e., stacked) analyses.

Finally, inverse Compton scattering of CMB photons by cluster electrons generates a small polarisation signal, yet to be detected. Its main contribution arises from the

CMB quadrupole at the cluster's location, which, in principle, could be used to measure the local CMB quadrupole at different cluster locations, thus circumventing cosmic variance (Kamionkowski and Loeb, 1997).

1.4 Galaxy cluster lensing of the cosmic microwave background

1.4.1 The cosmic microwave background

The cosmic microwave background (CMB) is the leftover electromagnetic radiation from the Big Bang. It was released at a redshift $z = 1089.80 \pm 0.21$ (Planck 2018 results VI, 2018), roughly 380 000 years after the Big Bang, at a time known as last scattering, when the Big Bang thermal photons decoupled from the baryons and were able to free-stream for the first time in cosmic history. Before decoupling, the CMB photons were tightly-coupled to baryons through Compton scattering and had a thermal (blackbody) distribution. After decoupling, they have undergone little scattering, retaining their thermal distribution. Indeed, the observed CMB has a blackbody spectrum to a very good approximation, with an average temperature $T_{\mathrm{CMB}} = 2.7255 \pm 0.0006\,\mathrm{K}$ (Mather et al., 1994).

The CMB is very isotropic across the sky, reflecting the homogeneity of the early Universe. However, small anisotropies exist [$\mathscr{O}(10^{-5})$], caused mostly by inhomogeneities in the last scattering surface (the progenitors of late-time LSS), in addition to smaller 'late time' contributions (Doroshkevich et al., 1978; Seljak, 1994; Hu and Sugiyama, 1995). The CMB is also partially linearly polarised, at about the level of 10 % of the temperature anisotropies, due to the presence of a quadrupole in the temperature angular distribution at last scattering and to the anisotropic nature of Thomson scattering (Rees, 1968; Bond and Efstathiou, 1984; Polnarev, 1985). The study of these temperature and polarisation anisotropies has revolutionised our knowledge of the Universe over the past three decades (see, e.g., Planck 2018 results I 2018).

First predicted in the late 1940s by Alpher and Herman (Alpher and Herman, 1948), who estimated its temperature to be of a few K, the CMB was serendipitously discovered by Penzias and Wilson about 20 years later (Penzias and Wilson, 1965), whose measurement of a background radiation were soon recognised to have a cosmic origin (Dicke et al., 1965; Peebles, 1968). It took about three decades for the temperature anisotropies to be detected, by the space-based COBE mission (Smoot et al., 1992). (The temperature dipole, mostly a Doppler shift due to our peculiar motion, was,

however, detected in the early 1970s.) Shortly after, the acoustic peaks were first found by a series of ground and balloon-based experiments (Miller et al., 1999; de Bernardis et al., 2000; Hanany et al., 2000), and so was the polarisation signal (Kovac et al., 2002). In the past two decades, the space-based WMAP and *Planck* experiments, as well as various ground-based experiments, most notably SPT and ACT, and their subsequent incarnations, have mapped the temperature and polarisation anisotropies with increasing sensitivity, establishing the ΛCDM paradigm as our current standard picture of the Universe (Hinshaw et al., 2013; Planck 2018 results VI, 2018).

The CMB anisotropies are a particularly robust cosmological probe, in part due to the fact that they can be accurately calculated with linear theory, since inhomogeneities at decoupling are very small (e.g., Seljak 1994; Hu and Dodelson 2002; Challinor 2004; Dodelson 2004). The primary anisotropies, those sourced at last scattering, arise mostly from variations in the photon density and gravitational potential across the last-scattering surface. There is also a smaller Doppler-shift contribution from the peculiar velocities at last scattering.

These primary anisotropies are statistically very close to Gaussian, with no detection of non-Gaussianity to date (e.g., Planck 2018 results VIII 2018). Thus, all of their information is contained in the angular power spectrum. Let $T(\hat{\mathbf{n}})$ be the CMB temperature at angular position $\hat{\mathbf{n}}$. Its difference with respect to the mean CMB temperature, $\Delta T(\hat{\mathbf{n}}) \equiv T(\hat{\mathbf{n}}) - T_{\rm CMB}$, can be expressed in terms of spherical harmonics,

$$\Delta T(\hat{\mathbf{n}}) = \sum_{l=0}^{\infty} \sum_{m=-l}^{m=l} a_{lm} Y_{lm}(\hat{\mathbf{n}}), \tag{1.41}$$

where $Y_{lm}(\hat{\mathbf{n}})$ are the spherical harmonics, and a_{lm} are the spherical-harmonic coefficients. The multipole l indicates the typical angular scale that the corresponding coefficients a_{lm} capture. Higher l corresponds to smaller angular scales, with a typical angular scale of variation $\theta \sim \pi/l$. The angular temperature, or TT, power spectrum is then

$$\langle a_{lm} a^*_{l'm'} \rangle = \delta_{ll'} \delta_{mm'} C_l^{TT}, \tag{1.42}$$

where the Kronecker delta functions are a result of statistical isotropy, as is the fact that C_l^{TT} depends only on l.

The CMB TT power spectrum as measured by the *Planck* mission is shown in Figure 1.8 as a function of l. The error bars denote $1\,\sigma$ uncertainties, whereas the blue curve is the best-fit prediction. Instead of C_l^{TT}, the standard quantity $\mathscr{D}_l^{TT} \equiv l(l+1)C_l^{TT}/2\pi$ is plotted. Three main features can be identified visually. At low l, $\mathscr{D}_l^{TT}$ is approximately

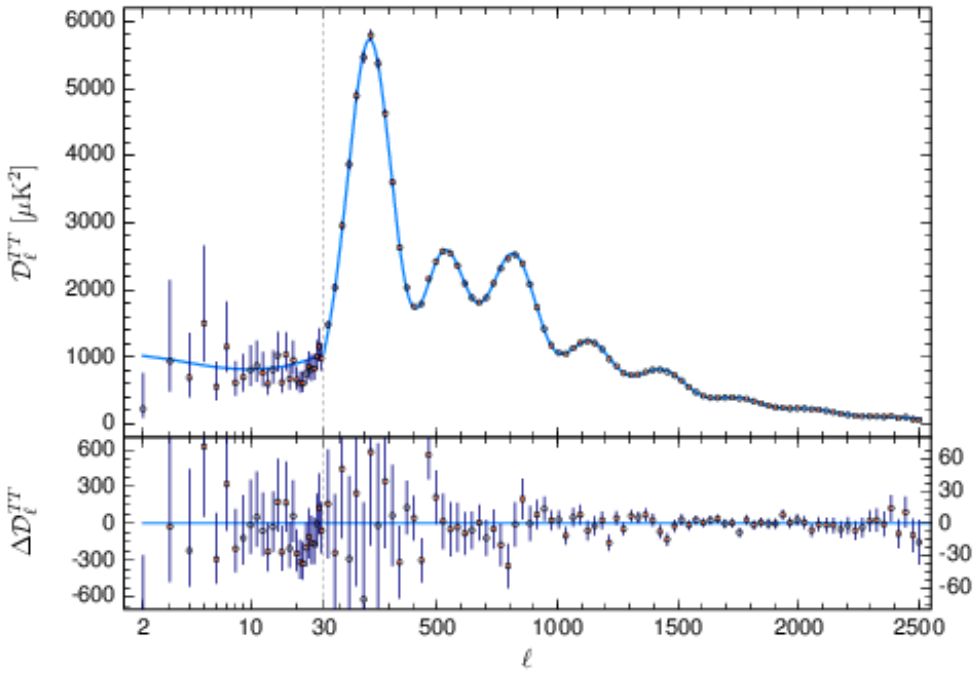


Fig. 1.8 *Planck* 2018 TT power spectrum. Figure credit: Planck 2018 results VI (2018).

constant. These angular scales correspond to physical scales that were outside the Hubble horizon at the time of last scattering, and as such are a direct probe of the primordial fluctuations. That $\mathscr{D}_l^{TT}$ is roughly constant on these large scales is a consequence of the primordial power spectrum being almost scale-invariant. An additional contribution on large scales comes from the decay of the gravitational potential at late times due to dark energy, which induces an additional gravitational shifting as photons travel from the last-scattering surface towards us (the 'integrated Sachs-Wolfe effect'; Sachs and Wolfe 1967). At higher l, the baryon acoustic oscillations are observed. As explained in Section 1.2.3, these oscillatory features are a result of sound waves propagating in the pre-decoupling baryon-photon plasma. These oscillations are damped on small scales due to photon diffusion, in what is known as Silk, or diffusion, damping (Silk, 1968). Before decoupling, on scales smaller than $\mathscr{O}(10)$ Mpc, the mean-free path for Thomson scattering is non-negligible and photons and baryons are not perfectly coupled. Photons can diffuse out of small-scale overdense regions, erasing small-scale inhomogeneities in the process. Further damping arises from the fact that decoupling did not happen instantaneously.

The (linear) CMB polarisation field can be described with two Stokes' parameters, $Q(\hat{\mathbf{n}})$ and $U(\hat{\mathbf{n}})$, which are directly observable, or, equivalently, with the E and B-mode

polarisation fields, $E(\hat{\mathbf{n}})$ and $B(\hat{\mathbf{n}})$, which can be obtained from the Q and U fields (see, e.g., Dodelson 2004). Unlike Q and U, the E and B-mode fields have the advantage of being coordinate invariant. In analogy with the TT case, the EE, TE, and BB power spectra can be detected and used as a source of cosmological information (the TB and EB power spectra vanish due to the fact that T and B, and E and B, have opposite parity). Scalar perturbations only generate E-modes to linear order. Primordial tensor perturbations, i.e., a stochastic cosmological background of primordial gravitational waves, however, source both E and B-modes (Seljak and Zaldarriaga, 1997): finding these primordially-sourced B-modes has become one of the holy grails of modern cosmology.

1.4.2 CMB lensing

Given the increasing sensitivity of current and future CMB experiments, there is at present a large interest in finding smaller, *secondary* signals in CMB observations, and in using them as cosmological probes. Some of these signals are generated as the CMB photons propagate from the last-scattering surface towards us; in this context, the CMB can be thought of as a 'backlight'. These include the various SZ effects discussed above, the ISW and Rees-Sciama effects (Rees and Sciama 1968; see, e.g., Cai et al. 2010), and CMB lensing.

As the CMB photons free-stream from the last scattering surface towards us, they are deflected as a result of the gravity of the matter distribution through which they travel. This effect is known as CMB lensing (Blanchard and Schneider 1987; Cole and Efstathiou 1989; see Lewis and Challinor 2006 for a review). First detected about a decade ago (Smith et al., 2007; Das et al., 2011), CMB lensing has now been detected to $40\,\sigma$ and has become a powerful cosmological probe (Planck 2018 results VIII, 2018).

Let X be an 'unlensed' CMB field, i.e., a CMB field as it would have been observed if there was no lensing, where X can be T, Q, or U. Lensing remaps the CMB fields so that the 'lensed' field $\tilde{X}(\hat{\mathbf{n}})$ along the line-of-sight direction $\hat{\mathbf{n}}$ is the unlensed field at $\hat{\mathbf{n}} + \boldsymbol{\alpha}(\hat{\mathbf{n}})$, i.e., $\tilde{X}(\hat{\mathbf{n}}) = X(\hat{\mathbf{n}} + \boldsymbol{\alpha})$.

The total deflection angle, $\boldsymbol{\alpha}$, can be seen as the addition of all the deflections due to all the matter that a given photon encounters along its way. Consider a deflection at a comoving distance χ from the observer, and let us assume that the CMB comes from a single source plane (the distance to last scattering is about 14 000 Mpc comoving, which is large compared to the width of the last-scattering surface, about 100 Mpc comoving, so this ought to be a good approximation, Lewis and Challinor 2006). The local deflection angle, $\delta\boldsymbol{\beta}$, can be written as $\delta\boldsymbol{\beta} = -2\boldsymbol{\nabla}_\perp \Psi \delta\chi$ (e.g., Lewis and Challinor

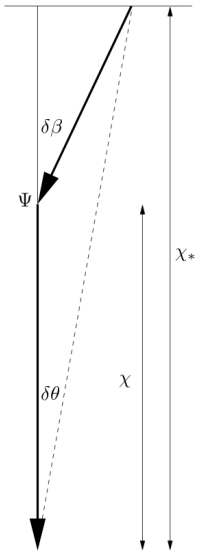


Fig. 1.9 CMB lensing geometry. Here, $\delta\beta$ denotes the local deflection angle due to Ψ, which gives rise to an observed deflection angle $\delta\theta$; χ and $\chi_\star$ denote, respectively, the comoving distances to the lens and to the source (the last scattering surface). Figure credit: Lewis and Challinor (2006).

2006), where Ψ is the Newtonian gravitational potential (or, in a general relativistic framework, the Weyl potential), $\delta\chi$ denotes a small comoving distance interval, and $\boldsymbol{\nabla}_\perp$ denotes the two-dimensional gradient perpendicular to the direction of travel, along which $\delta\chi$ runs. For simplicity, let us assume that the Universe is spatially flat. Following a simple geometric argument (see Figure 1.9), this local deflection angle translates into an observed deflection angle

$$\delta\boldsymbol{\theta} = \frac{\chi_\star - \chi}{\chi_\star}\delta\boldsymbol{\beta} = -2\frac{\chi_\star - \chi}{\chi_\star}\boldsymbol{\nabla}_\perp\Psi\delta\chi = -2\frac{\chi_\star - \chi}{\chi_\star\chi}\boldsymbol{\nabla}_{\hat{\mathbf{n}}}\Psi\delta\chi, \tag{1.43}$$

where $\chi_\star$ is the comoving distance to the source, i.e., to last scattering, and where $\boldsymbol{\nabla}_{\hat{\mathbf{n}}}$ denotes the gradient with respect to the observation angular coordinates.

In the limit in which the deflection angles $\delta\boldsymbol{\beta}$ and their angular derivatives are small, known as the 'weak lensing' limit, the total deflection angle, $\boldsymbol{\alpha}$, can be computed by adding all the deflections evaluated along the undeflected line of sight (the Born approximation; see Lewis and Challinor 2006). That is,

$$\boldsymbol{\alpha} = -2\boldsymbol{\nabla}_{\hat{\mathbf{n}}}\int_0^{\chi_\star}\frac{\chi_\star - \chi}{\chi_\star\chi}\Psi(\chi\hat{\mathbf{n}}; z(\chi))d\chi \equiv \boldsymbol{\nabla}_{\hat{\mathbf{n}}}\psi, \tag{1.44}$$

where Ψ is evaluated at the appropriate location, $\chi\hat{\mathbf{n}}$, and redshift, $z(\chi)$. The *lensing potential*, ψ, has been defined such that $\boldsymbol{\alpha} = \boldsymbol{\nabla}_{\hat{\mathbf{n}}}\psi$. It can be seen as an integral of

the gravitational potential along the line of sight all the way back to last scattering, weighted by the *lensing kernel*.

An important quantity to consider is the Jacobian of the deflected directions, which contains information about how the sizes and shapes of the anisotropies are distorted. To linear order, it can be written as (Lewis and Challinor, 2006)

$$A_{ij} \equiv \delta_{ij} + \nabla_i \alpha_j = \begin{pmatrix} 1-\kappa-\gamma_1 & -\gamma_2 \\ -\gamma_2 & 1-\kappa+\gamma_1 \end{pmatrix}. \tag{1.45}$$

Here, κ is the *lensing convergence*, which is responsible for the trace of the distortion, i.e., it quantifies the apparent local size change due to lensing. On the other hand, γ_1 and γ_2 are the two components of the *lensing shear*, which are responsible for the traceless distortion, i.e., they quantify the apparent area-preserving local shape change. Both quantities can be obtained from the lensing potential, and thus contain the same information. In particular, the convergence can be written as $\kappa = -\frac{1}{2}\nabla^2_{\hat{\mathbf{n}}}\psi$, which is a projected version of Poisson's equation. By using the definition of ψ (see Eq. 1.44) in this equation, and then using Poisson's equation, it can be seen that the lensing convergence is a weighted integral of the matter distribution along the line of sight.

CMB lensing is a small effect, with associated changes of the order of $10\,\mu$K in the observed CMB temperature. The deflections are typically small, of about a few arcmin for linear LSS and for galaxy clusters. Compact objects such as black holes can produce much larger deflection angles, but only a small fraction of lines of sight pass near them, so they can be ignored. The weak lensing approximation is thus expected to be a good approximation, at least for current observations (Lewis and Challinor, 2006). However, despite being small, the lensing deflections are correlated over larger scales. For linear structure, the main source of CMB lensing, a typical potential has a size of around 300 Mpc comoving (about the peak of the matter power spectrum). If it is half-way back to the last scattering surface, $\chi \sim 7\,000$ Mpc, its observed size is around 2° (Lewis and Challinor, 2006). The CMB lensing deflections are thus correlated over degree scales, which means that CMB observables are expected to be affected by lensing on these scales.

Indeed, the TT, EE, and TE power spectra are modified by lensing on degree scales due to lensing by LSS. In particular, the acoustic peaks are smoothed at a several percent level (Lewis and Challinor, 2006). Intuitively, lensing, by distorting the sizes and shapes of the hot and cold spots of the CMB anisotropies, 'blurs' the BAO scale, which translates into smoothing of the BAO oscillations in harmonic space. The total variance, however, is conserved, as lensing only remaps the anisotropies: lensing drives

power towards smaller scales, where power is increased. For the TT power spectrum, at $l > 2000$, where there is little unlensed power, lensing is in fact an order unity effect (Lewis and Challinor, 2006).

In addition, lensing of the E-mode polarisation field generates a B-mode polarisation field with an approximately white noise power spectrum on large scales, even if no primordial B-modes are present (Zaldarriaga and Seljak, 1998). These lensing B-modes, whose detection was first reported in Hanson et al. (2013), have now been detected to $18.1\,\sigma$ (Sayre et al., 2019), and currently constitute the main limitation for the detection of primordial, inflation-sourced B-modes. To look optimally for primordial B-modes, the CMB can be 'delensed' by remapping back the particular realisation of the anisotropies that we see with an estimate of the lensing potential, thus circumventing the B-mode sample variance due to lensing (e.g., Seljak and Hirata 2004).

Lensing also generates a non-Gaussian signal in the CMB anisotropies. In effect, even if the unlensed CMB and the lensing potential are assumed to be Gaussian, remapping of a Gaussian field by a Gaussian field results in a non-Gaussian field. This lensing-induced non-Gaussianity can be probed by measuring the three-point, and higher-order, correlation functions that it induces (or, equivalently, the bispectrum, trispectrum, etc. in harmonic space); see, e.g., Lewis and Challinor (2006).

Finally, it should be noted that lensing, as a purely geometric effect, does not alter the CMB frequency spectrum. Thus, multi-frequency observations cannot be used in order to separate the lensing signal from the unlensed CMB anisotropies, as is the case, e.g., of the tSZ effect. However, lensing can be probed at the map level thanks to our very good statistical knowledge of the angular distribution of the unlensed anisotropies, as we will see in the next section.

1.4.3 CMB lensing reconstruction

The lensing potential (or, equivalently, the deflection field, or the convergence) can be estimated, or 'reconstructed', with CMB observations alone due to the specific correlations lensing induces in the CMB fields.

The simplest technique for CMB lensing reconstruction involves CMB lensing quadratic estimators, so called because they are quadratic in the CMB fields (Hu, 2001; Hu and Okamoto, 2002). Here, we present the TT quadratic estimator, which takes as input two copies of the observed temperature field, following the discussion of Lewis and Challinor (2006). For simplicity, we will work in the flat-sky approximation, which is valid if the observed fields are small enough such that the curvature of the sky can be neglected. Within this approximation, Fourier transforms can be used instead of

spherical harmonics, greatly simplifying calculations. An analogous estimator can be derived for the spherical sky (see Okamoto and Hu 2003).

Consider the lensed CMB temperature, $\tilde{T}(\hat{\mathbf{n}}) = T(\hat{\mathbf{n}} + \boldsymbol{\nabla}_{\hat{\mathbf{n}}}\psi)$. It can be expanded to linear order in the lensing potential,

$$\tilde{T}(\hat{\mathbf{n}}) = T(\hat{\mathbf{n}} + \boldsymbol{\nabla}_{\hat{\mathbf{n}}}\psi) \simeq T(\hat{\mathbf{n}}) + \boldsymbol{\nabla}_{\hat{\mathbf{n}}}\psi(\hat{\mathbf{n}}) \cdot \nabla_{\hat{\mathbf{n}}} T(\hat{\mathbf{n}}), \tag{1.46}$$

which in Fourier space reads

$$\tilde{T}(\mathbf{l}) \simeq T(\mathbf{l}) - \int \frac{d^2\mathbf{l}'}{2\pi} \mathbf{l}' \cdot (\mathbf{l} - \mathbf{l}')\psi(\mathbf{l} - \mathbf{l}')T(\mathbf{l}'). \tag{1.47}$$

This 'gradient approximation' is valid on scales on which deflections are small compared to the wavelength of the unlensed fields, roughly for $l < 2000$. However, on very small scales, $l > 3000$, this approximation is also valid, since due to damping the unlensed CMB is very smooth on these scales, essentially just a gradient. Under this approximation, the correlation between two different modes of the lensed temperature field can be written to linear order in ψ as

$$\left\langle \tilde{T}(\mathbf{l})\tilde{T}^*(\mathbf{l} - \mathbf{L}) \right\rangle_T = \delta(\mathbf{L})C_l^{TT} + \frac{1}{2\pi}\left[(\mathbf{L} - \mathbf{l}) \cdot \mathbf{L}C_{|\mathbf{l}-\mathbf{L}|}^{TT} + \mathbf{l} \cdot \mathbf{L}C_l^{TT}\right]\psi(\mathbf{L}), \tag{1.48}$$

where C_l^{TT} is the unlensed TT power spectrum, and where the angular brackets denote ensemble averaging over the unlensed temperature for fixed lensing potential. Here, the first term on the left-hand side is the isotropic unlensed power, and the second term corresponds to the lensing-induced off-diagonal correlations, which are proportional to the lensing potential, $\psi(\mathbf{L})$. Thus, these off-diagonal correlations can be used in order to construct an estimator of $\psi(\mathbf{L})$, $\hat{\psi}(\mathbf{L})$, by summing over all possible combinations of the quadratic product $\tilde{T}(\mathbf{l})\tilde{T}^*(\mathbf{l} - \mathbf{L})$ with $\mathbf{L} \neq \mathbf{0}$,

$$\hat{\psi}(\mathbf{L}) = N(\mathbf{L}) \int \frac{d^2\mathbf{l}}{2\pi} \hat{T}(\mathbf{l})\hat{T}\,(\mathbf{l} - \mathbf{L})\, g(\mathbf{l}, \mathbf{L}), \tag{1.49}$$

where $\hat{T}$ denotes a measurement (estimate) of the lensed temperature, $\tilde{T}$, $g(\mathbf{l}, \mathbf{L})$ is a weight function chosen to ensure optimality, and $N(\mathbf{L})$ is a normalisation factor. This is the TT quadratic estimator. Although here it is presented as an estimator of the lensing potential, estimators for the deflection field, the shear, and the convergence can be trivially obtained, as these quantities are just derivatives of the lensing potential; e.g., the quadratic estimator of the lensing convergence is $\hat{\kappa}(\mathbf{L}) = L^2\hat{\psi}(\mathbf{L})/2$.

By requiring that $\hat{\psi}(\mathbf{L})$ is unbiased when ensemble averaging over the unlensed temperature, $\left\langle \hat{\psi}(\mathbf{L}) \right\rangle_T = \psi(\mathbf{L})$, $N(\mathbf{L})$ is

$$N(\mathbf{L}) = \left[\int \frac{d^2\mathbf{l}}{(2\pi)^2} \left[(\mathbf{L}-\mathbf{l}) \cdot \mathbf{L} C^{TT}_{|\mathbf{l}-\mathbf{L}|} + \mathbf{l} \cdot \mathbf{L} C^{TT}_{l} \right] g(\mathbf{l}, \mathbf{L}) \right]^{-1}. \tag{1.50}$$

By further requiring that the estimator has minimum variance (to linear order in ψ), the weight function can be written as (Hu, 2001)

$$g(\mathbf{l}, \mathbf{L}) = \frac{(\mathbf{L}-\mathbf{l}) \cdot \mathbf{L} C^{TT}_{|\mathbf{l}-\mathbf{L}|} + \mathbf{l} \cdot \mathbf{L} C^{TT}_{l}}{2 C^{\hat{T}\hat{T}}_{l} C^{\hat{T}\hat{T}}_{|\mathbf{l}-\mathbf{L}|}}, \tag{1.51}$$

where $C^{\hat{T}\hat{T}}_{l}$ is the total power spectrum of the observed temperature, which, to linear order in ψ, can be written as $C^{\hat{T}\hat{T}}_{l} = \tilde{C}^{\tilde{T}\tilde{T}}_{l} + N^{TT}_{l}$, where $\tilde{C}^{\tilde{T}\tilde{T}}_{l}$ is the lensed TT power spectrum, and where N^{TT}_{l} is the beam-deconvolved temperature noise power spectrum. This guarantees that the TT quadratic estimator has minimum variance, although it does not imply that this is the optimal temperature-based lensing estimator.

With this choice of weight function, $N(\mathbf{L})$ is also the estimator variance to linear order in ψ. As is apparent from Eq. (1.50), this variance arises not only from instrumental noise, but also from the unlensed temperature fluctuations themselves, which limit how well the lensing potential can be estimated.

This TT quadratic estimator can be derived from a maximum-likelihood approach in which the likelihood of the observed temperature given the lensing potential is assumed to be Gaussian, arising as the first step in a Newton-Raphson-like approach to maximise the likelihood (see Hirata and Seljak 2003 and Hanson et al. 2010). Moreover, it can be understood in a more intuitive way in the 'large-scale lens regime', in which small-scale (roughly $l > 300$) temperature modes are used in order to reconstruct large scale (roughly $l < 300$) lensing modes (see Schaan and Ferraro 2019 and Fabbian et al. 2019). In this regime, the TT quadratic estimator can be expressed as a sum of a magnification-only estimator and a shear-only estimator. Intuitively, the magnification-only estimator reconstructs the lensing potential by comparing the local size of the hot and cold spots in the CMB anisotropies with their average size as expected from the unlensed TT power spectrum. For example, if in a region the anisotropies are larger than average, a positive convergence (i.e., an overdensity) is predicted. Similarly, the shear-only estimator looks for local deviations from statistical isotropy in order to reconstruct the potential. Thus, the quadratic estimator can be understood as using both information from the local magnification and shearing of the CMB anisotropies. It should be noted that both the magnification and shear-only parts are unbiased on all

scales, although their interpretation as convergence and shear-only estimators is only clear in the large-scale lens regime. They can also be used independently from each other. In fact, it is argued in Schaan and Ferraro (2019) that the shear-only estimator is essentially insensitive to uncleaned foregrounds, such as the difficult-to-remove kSZ signal, which makes it a very interesting tool, especially for future experiments, which will probe high multipoles ($l > 2000$), where the kSZ signal is expected to be relevant.

Similar quadratic estimators can be derived for all possible combinations of the CMB temperature and the polarisation fields, TT, TE, TB, EE, and EB (see Hu and Okamoto 2002). In addition, they can all be jointly combined into a minimum-variance quadratic estimator. Their associated reconstruction noise per mode ($N(\mathbf{L})$ in Eq. 1.50 for the TT estimator, with analogous expressions for the other estimators) for the *Planck* 2015 lensing analysis are shown, as a function of lensing multipole, L, in Figure 1.10. It can be seen how, for *Planck*, the TT estimator dominates the total signal-to-noise, with a significant contribution from polarisation only on large scales. The situation, however, will be different in future experiments, in which polarisation will make a significant part, if not the most important part, of the total signal-to-noise (see, e.g., Abazajian et al. 2016 and Simons Observatory Collaboration 2019). Furthermore, polarisation also provides different sensitivity to foreground contamination, being in general more robust against biases from uncleaned foregrounds. For instance, whereas the kSZ effect is a difficult-to-remove source of bias in the temperature at high l, the polarised total SZ signal is very small, about 10–100 nK (Carlstrom et al., 2011), so polarisation estimators are essentially unaffected by it.

Due to their simplicity and their computational efficiency, quadratic estimators have been widely used in lensing analyses. Moreover, for the resolution and noise levels of experiments like *Planck*, they are very close to optimal. In Chapter 2, for instance, we use the TT quadratic estimator, which is shown in its computationally efficient form in Eq. (2.3), in order to estimate the lensing convergence around the sky locations of the *Planck* galaxy clusters. However, for future experiments, the quadratic estimators are suboptimal. More optimal approaches, in which the full lensed CMB likelihood (or posterior) is maximised, or explored, have been developed (see Anderes et al. 2015, Carron and Lewis 2017, and Millea et al. 2019).

CMB lensing reconstruction allows one to obtain the lensing convergence power spectrum, $C_l^{\kappa\kappa}$, from which constraints on parameters such as Ω_m and σ_8 can be derived (e.g., Planck 2018 results VIII 2018). In addition, reconstructed lensing maps can also be used for delensing (e.g., Carron et al. 2017), and for cross-correlation with other tracers of the cosmic density field, such as galaxies (e.g., Bianchini et al. 2015),

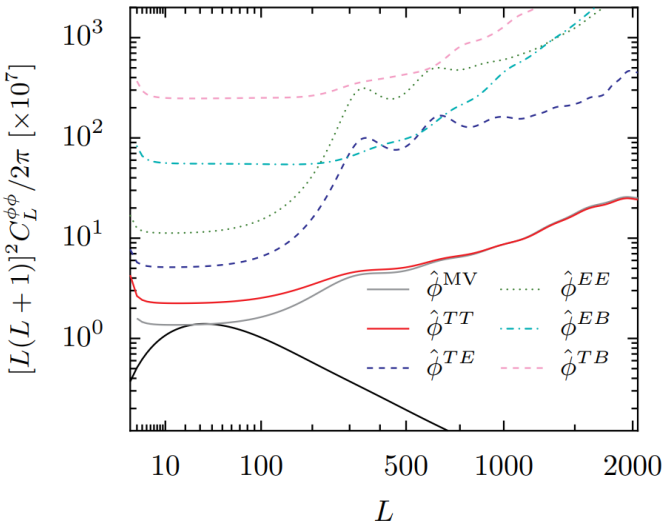


Fig. 1.10 CMB lensing quadratic estimator reconstruction noise as a function of lensing multipole, L, for the different quadratic estimators and their minimum-variance combination for the *Planck* 2015 lensing analysis. In addition, the lensing potential power spectrum is shown in black. Figure credit: Planck 2015 results XV (2016).

quasars (e.g., Hirata et al. 2008), the tSZ signal across the sky (e.g., Hill and Spergel 2014), the X-ray signal across the sky (e.g., Hurier et al. 2019), and the cosmic infrared background (e.g., Planck 2013 results XVIII 2014).

1.4.4 Galaxy cluster masses from CMB lensing

Galaxy clusters leave a specific CMB lensing signature, which can be used in order to estimate their masses. A typical cluster produces CMB deflections of the order of a few arcmin around its location in the sky. On arcmin scales, the unlensed CMB is roughly a gradient, and therefore Eq. (1.46) is accurate. If the cluster lensing potential is roughly circularly symmetric, then, locally, the CMB as lensed by a cluster is a dipole centred at the cluster location and aligned with the direction of the unlensed gradient, oriented such that the amplitude of the unlensed gradient is decreased (Seljak and Zaldarriaga, 2000). The magnitude of this effect is proportional to both the unlensed gradient and the deflection field, and so depends on the cluster mass. The r.m.s. temperature gradient is about $10\,\mu\mathrm{K\,arcmin}^{-1}$, and so for typical deflections of about 1 arcmin, cluster lensing produces variations of about $10\,\mu$K. For polarisation, gradients are of about $1\,\mu\mathrm{K\,arcmin}^{-1}$, and therefore a lensing signal of about $1\,\mu$K is generated (Lewis and Challinor, 2006).

Current CMB experiments are not sensitive enough to enable the measurement of individual cluster masses to high signal-to-noise, as opposed to other techniques, such as galaxy weak lensing. However, ensemble-averaged cluster masses can be obtained to relatively high significance, which can prove very useful in the context of counts analyses, as they do not depend on the clusters' dynamical state (see, e.g., Planck 2015 results XXIV 2016, and Chapters 2 and 3). CMB lensing cluster mass determination has other interesting intrinsic virtues. The signal-to-noise does not decrease significantly with redshift (e.g., Melin and Bartlett 2015), and it does not rely on there being a high density of background galaxies, an important limitation of galaxy weak lensing mass estimation at high redshift. This makes CMB lensing mass estimation especially useful in this regime. Moreover, CMB lensing masses and galaxy weak-lensing masses are affected by different systematic effects, which makes CMB lensing valuable at low redshift too. In particular, it does not suffer from galaxy misidentification and from photometric redshift errors. However, it is sensitive to contamination from residual tSZ signals (a cluster's tSZ signal can be up to an order of magnitude larger than the lensing signal, although it can be in principle subtracted out; see, e.g., Melin and Bartlett 2015; Patil et al. 2020), and, most importantly, from the kSZ signal (e.g., Melin and Bartlett 2015; Raghunathan et al. 2017, 2019a). Other potential sources of bias include miscentering of the cluster profile used to model the lensing signal, and deviations of this profile from the true cluster profile (see, e.g., Raghunathan et al. 2017, and Chapters 2, 3, and 4).

Several approaches have been proposed in order to estimate cluster masses from CMB lensing. One possibility is to reconstruct the lensing convergence in a non-parametric way, using the quadratic estimators or any other technique, and then to fit a (mass-dependent) cluster convergence model, as proposed, e.g., in Hu et al. (2007) and Melin and Bartlett (2015). This was the approach followed in Planck 2015 results XXIV (2016), in which the method presented in Melin and Bartlett (2015), using a TT quadratic estimator and matched-filtering with a convergence model, was applied to estimate the average mass of 433 tSZ-detected *Planck* clusters, obtaining a $5\,\sigma$ detection of cluster CMB lensing. Other works in which the cluster convergence is estimated non-parametrically include Baxter et al. (2017), in which the CMB lensing signal of 3 697 clusters optically-selected from the Dark Energy Survey (DES) is detected to $8.1\,\sigma$ significance from SPT temperature data, and Geach and Peacock (2017), in which the masses of 26 111 optically-selected SDSS clusters, estimated from *Planck* convergence maps, are used in order to calibrate their mass-richness relation at the 10 %

level. We also follow an approach in which the cluster convergence is reconstructed non-parametrically in Chapter 2.

Other approaches do not include an intermediate CMB lensing reconstruction step, but model the cluster CMB lensing signal directly. These include maximum-likelihood methods, such as those proposed in Lewis and King (2006), Yoo and Zaldarriaga (2008), and Yoo et al. (2010), and the method applied in Baxter et al. (2015), in which the CMB lensing signal of 513 tSZ-detected SPT clusters is detected to $3.1\,\sigma$ significance using SPT temperature data. Parametric Bayesian methods have also been proposed (e.g., Raghunathan et al. 2017). Although quadratic-estimator-based methods are nearly optimal for experiments such as *Planck*, maximum-likelihood and Bayesian methods will outperform them in future experiments such as CMB-S4 (see, e.g., Raghunathan et al. 2017).

Methods in which the unlensed CMB gradient is measured and 'factored out' of the high-pass-filtered observed temperature field in order to measure directly the cluster deflection field have also been proposed (Seljak and Zaldarriaga 2000; Vale et al. 2004; Horowitz et al. 2019), and are also expected to outperform quadratic-estimator-based methods in the future (see, e.g., Horowitz et al. 2019). They rely on the fact that the unlensed CMB gradient has little contribution from scales smaller than $l \sim 2000$, and that cluster lensing barely modifies this gradient on scales $l < 2000$, so it can be in principle measured directly from observations and 'fitted out'. A method along these lines was applied in Raghunathan et al. (2019c) to SPTpol polarisation maps, detecting the CMB lensing signal of 18 000 DES clusters to $4.8\,\sigma$, the first detection of polarisation-only-based CMB lensing by clusters.

CMB lensing by halos smaller than galaxy clusters has also been detected, with Madhavacheril et al. (2015) reporting a $3.2\,\sigma$ detection of the CMB lensing signal of 12 000 optically-selected CMASS galaxies from the SDSS-III/BOSS survey with ACTPol temperature data, the first detection of CMB lensing by compact objects. There have also been recent first detections of CMB lensing by filaments (He et al., 2018) and by voids (Cai et al., 2017; Raghunathan et al., 2020).

1.5 Current cosmological constraints

Cluster counts became a precision cosmological probe in the first decade of the 21st century, mostly through the analysis of X-ray cluster samples from ROSAT catalogues (Borgani et al., 2001; Reiprich and Böhringer, 2002; Seljak, 2002; Viana et al., 2002; Allen et al., 2003; Pierpaoli et al., 2003; Schuecker et al., 2003; Voevodkin and Vikhlinin,

2004; Henry, 2004; Mantz et al., 2008; Vikhlinin et al., 2009; Henry et al., 2009; Mantz et al., 2010), but also thanks to optical analyses (Rozo et al., 2010). These works used widely different cluster samples and a range of observables such as the X-ray luminosity, the most commonly employed, the gas mass, Y_X (X-ray), and richness (optical). They consistently found values of Ω_m around 0.3 and of σ_8 around 0.8, in agreement with the constraints derived from CMB experiments such as WMAP (Hinshaw et al., 2013). This is a remarkable agreement given the huge redshift difference between galaxy clusters and the CMB, the very different physics that come into play in these two different cosmological probes, and the very different systematic effects to which they are sensitive.

Over the past 10 years there has been significant further progress in both X-ray (Mantz et al., 2015; Schellenberger and Reiprich, 2017; Pacaud et al., 2018) and optical (Costanzi et al., 2019; Kirby et al., 2019) counts analyses, from which competitive cosmological constraints have been derived. These have been joined by the first generation of tSZ-based analyses (Benson et al., 2013; Hasselfield et al., 2013; Planck 2013 results XX, 2014; de Haan et al., 2016; Planck 2015 results XXIV, 2016; Bocquet et al., 2019). An important contribution to this progress has come from the now-widespread use of weak-lensing masses to anchor the absolute cluster mass scale. As the analyses of the previous decade, these studies differ widely in their methodology and in the cluster samples and mass observables employed, and they are also sensitive to different systematic effects. Nonetheless, the cosmological constraints derived from them are in excellent agreement with each other and with constraints derived from other low-redshift cosmological probes, such as cosmic shear. There is also good agreement with the constraints derived from the *Planck* CMB assuming ΛCDM, except perhaps in one parameter, σ_8. Indeed, recent low-redshift analyses, cluster counts included, generally find constraints on σ_8 that are somewhat lower than the *Planck*-derived one, $\sigma_8 = 0.8111 \pm 0.0060$ (*Planck* 2018 TT,TE,EE+lowE+lensing results; Planck 2018 results VI 2018).

This situation is illustrated in Figure 1.11, which shows the constraints on σ_8 at $\Omega_m = 0.3$ derived from a number of recent cosmological analyses. The constraints from cluster counts, which are classified into X-ray, SZ, and optical types, are shown in blue; those from cosmic shear, also a low-redshift probe, are shown in green for comparison; and the CMB-derived constraints are shown in red. Error bars denote $1\,\sigma$ uncertainties. It can be seen how all the low-redshift constraints on σ_8 are in very good agreement with each other. Moreover, as noted in Pratt et al. (2019), the empirical standard deviation of the cluster constraints around their empirical mean, shown as a light blue band, is

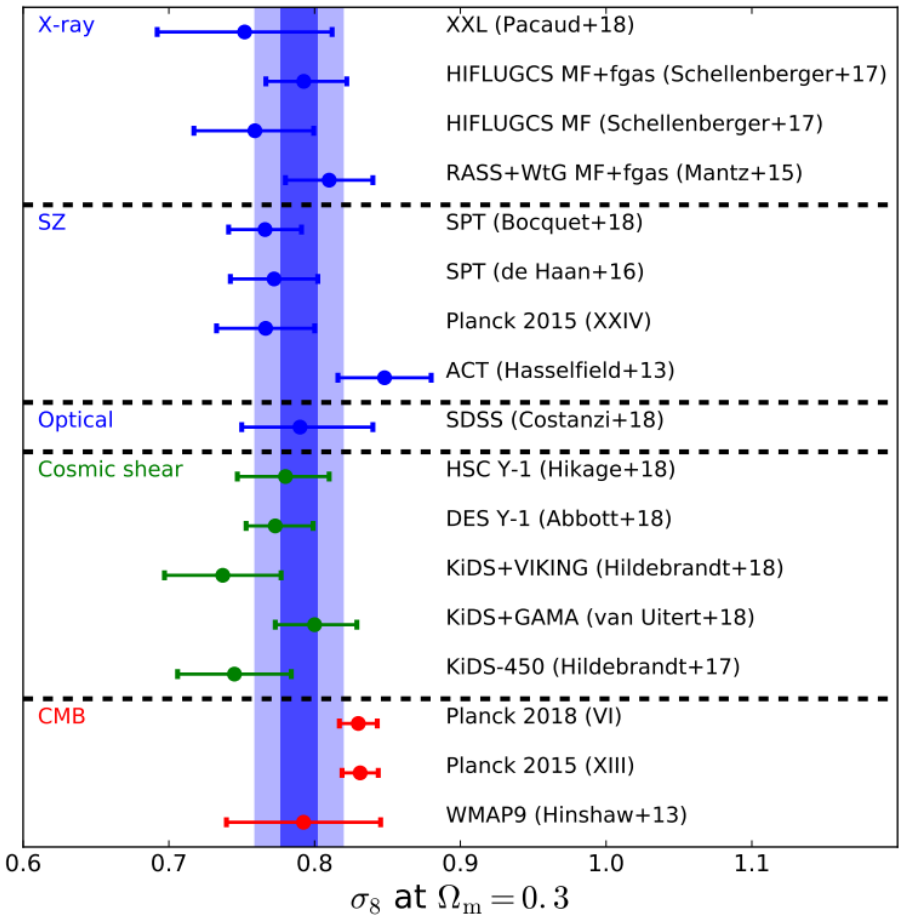


Fig. 1.11 Constraints on σ_8 at $\Omega_m = 0.3$ from a number of recent studies. Constraints from cluster counts (sometimes combined with gas fraction) are shown in blue, constraints from cosmic shear are shown in green, and CMB-derived constraints are shown in red. Error bars denote $1\,\sigma$ uncertainties. The light blue band corresponds to the standard deviation around the unweighted empirical mean of the seven independent cluster constraints, whereas the dark blue band corresponds to the empirical error on this unweighted mean. Figure credit: Pratt et al. 2019.

similar to the typical uncertainty of the individual cluster constraints, which indicates that uncertainties are well-understood and that confirmation bias is small. These low redshift constraints are also in good agreement with the constraint from WMAP, and, individually, with the much more precise constraint from *Planck*. However, the fact that all the cluster constraints, and indeed all the low-redshift constraints, except that from ACT (Hasselfield et al., 2013), are lower than the *Planck* constraint has attracted some attention in the community.

Unaccounted-for systematics in the less-robust low-redshift probes (as opposed to the better-understood CMB) are the most probable cause of this mild statistical discrepancy, at least from an Occam's razor point of view. For cluster counts, some likely candidates are absolute calibration errors of X-ray instruments, which can also affect X-ray-calibrated SZ analyses (temperatures measured by XMM-*Newton*, for instance, have been found to be systematically higher than those measured by *Chandra*; see, e.g., Schellenberger et al. 2015); relevant unaccounted-for relativistic tSZ corrections (e.g., Remazeilles et al. 2019); errors in the modelling of the ICM (e.g., Ruppin et al. 2019); and badly-understood completeness, due to clusters that should be detected being missed by detection algorithms (e.g., Xu et al. 2018). These systematics can be generally encapsulated as biases in the mass calibration of the cluster observables, which are in turn sometimes accounted for by an overall 'hydrostatic mass bias' parameter. Some recent studies have found this bias to be too large to be consistent with the range that is expected from deviations from hydrostatic equilibrium and non-thermal pressure support as observed in hydrodynamical simulations (e.g., Planck 2015 results XXIV 2016, where a bias of about 40 % is required to reconcile fully the *Planck* SZ counts with the *Planck* CMB; see also Chapter 3). This can therefore hint at further unaccounted-for systematics.

Alternatively, this apparent discrepancy could be the first sign of some unknown new physics that suppresses structure formation at low redshifts. This late-time suppression could be potentially achieved with, e.g., dark energy with $w > -1$, as it starts dominating earlier and therefore structure growth stalls earlier, and with more massive neutrinos, as they erase inhomogeneities through free-streaming. These proposed extensions to ΛCDM, however, seem not to be favoured by current data (see, e.g., Planck 2015 results XXIV 2016 and Bolliet et al. 2019).

1.6 Outline of the rest of the thesis

Galaxy cluster cosmology is therefore in a very exciting time: not only can clusters provide constraints that are currently competitive with those derived from other standard cosmological probes, but there is also a small apparent statistical discrepancy with high-redshift probes like the CMB to be addressed. In coming years, the quest to understand this apparent tension better will surely enhance our knowledge of clusters as cosmological probes and as astrophysical objects in their own right, and, if we get lucky, will perhaps also reveal something new about the Universe.

This thesis can be seen as a small contribution towards this aim. In the *Planck* 2015 SZ counts analysis, published in Planck 2015 results XXIV (2016), three different lensing-derived absolute mass calibrations were used for the baseline results. Two of them were obtained with galaxy weak-lensing masses (from CCCP and WtG), whereas the third calibration was obtained from CMB-lensing-derived masses, the first such example in the literature. Whereas the constraints on σ_8 for the two weak-lensing calibrations are in agreement with the *Planck* CMB (although, as usual, a little low), the constraint for the CMB lensing calibration, $\sigma_8 = 0.71 \pm 0.03$, is at a certain tension with it (see Figure 3.5 in Chapter 3).

We revisit this analysis for the particular case of the CMB lensing mass calibration. First, in Chapter 2 we measure from *Planck* temperature data the CMB lensing signal of most of the clusters in the sample that was used in the *Planck* cosmological analysis, the MMF3 cosmology sample, obtaining measurements that are independent from those used in Planck 2015 results XXIV (2016). Then, in Chapter 3 we perform a likelihood analysis of the MMF3 cosmology sample, taking as data the CMB lensing signal, tSZ signal, and redshift of each cluster in the sample. We employ a novel likelihood that does not require binning in data space and that does not marginalise over sky position, and which accounts for the selection of the cluster sample in a consistent way. Assuming a flat ΛCDM cosmology, we find cosmological constraints that are in agreement with those derived from the *Planck* CMB, including that on σ_8. Then, in Chapter 4 we calibrate the scatter and bias of our CMB lensing mass observable with mock observations from an N-body simulation, also quantifying deviations from log-normality. This calibration motivates the priors on the CMB lensing bias and scatter parameters that are used in Chapter 3. Finally, we present our conclusions in Chapter 5, where we also discuss briefly the near future of cluster cosmology and of CMB lensing cluster mass estimation.

Chapter 2

Estimation of the *Planck* cluster masses through CMB lensing

2.1 Introduction

As explained in Chapter 1, CMB photons coming from the last scattering surface are deflected due to the effect of the gravity of the matter distribution through which they travel, the observed net effect being a remapping of the CMB fluctuations on the sky by some deflection field $\boldsymbol{\alpha}(\hat{\boldsymbol{n}})$. This phenomenon, known as CMB lensing, induces additional correlations in the CMB anisotropies that make it possible to estimate, or reconstruct, the deflection field (or, equivalently, the lensing potential, or the convergence) with CMB observations. This allows to measure the masses of galaxy clusters through their CMB lensing signature: if the lensing convergence due to a galaxy cluster is reconstructed and a cluster convergence model is assumed, a cluster mass estimate can be obtained.

In this chapter we estimate the masses (or, more precisely, the signal-to-noise ratios) of the 433 galaxy clusters of the *Planck* CMB lensing calibration sample using *Planck* temperature data. These clusters are the clusters in the *Planck* MMF3 cosmology sample, which consists of a total of 439 SZ-detected clusters (see Planck 2015 results XXIV 2016 and Chapter 3), with measured redshift. Our procedure is similar to that followed in the CMB lensing cluster mass calibration presented in Planck 2015 results XXIV (2016), but there are likely minor differences in implementation. (Planck 2015 results XXIV 2016 do not provide full details of their implementation.) We then use these measured signal-to-noise values in the cosmological analysis of Chapter 3, in which they set the absolute cluster mass scale.

This chapter is structured as follows. First, in Section 2.2 we detail how we obtain clean temperature maps around the sky location of each cluster in our sample. Next, in Section 2.3 we describe the method that we follow in order to estimate the lensing convergence from these clean temperature maps. In Section 2.4 we describe our cluster convergence model and our mass observable, and we conclude in Section 2.5.

2.2 Production of clean temperature maps

2.2.1 Cluster fields, masks and apodisation

We use the six highest frequency (100–857 GHz) Full-Mission 2015 *Planck* temperature maps, from which we cut square patches of side 256 arcmin centred on the positions of the 433 clusters of the *Planck* CMB lensing calibration sample. We will refer to these cutouts as the 'cluster frequency fields'. Since the *Planck* maps are in the HEALPix pixelisation (Górski et al., 2005), we extract the cluster frequency fields as Cartesian projections centred at the cluster positions, setting the number of pixels of each field to be 512×512, with a pixel size of 0.5 arcmin. For the cluster central positions we use the Galactic coordinates estimated by the *Planck* collaboration through their cluster detection pipeline, which are publicly available in the PLA. We only use temperature maps, given that, for an experiment with the resolution and noise levels of *Planck*, the TT lensing quadratic estimator offers the best performance in terms of signal-to-noise (Hu and Okamoto, 2002).

We cosine apodise the edges of each cluster frequency field over a width of 12.8 arcmin to suppress spectral leakage when taking Fourier transforms. We find, using simulated CMB observations, that with such apodisation applied to a flat-sky, non-periodic CMB realisation with resolution and noise levels similar to *Planck*, we can recover the CMB power spectrum after simple scaling to account for the apodised mask. We also mask point sources using the union of the *Planck* 100 GHz and 143 GHz 2015 temperature point source masks provided in the PLA, and inpaint the masked regions with the method presented in Gruetjen et al. (2017) to reduce spectral leakage in later stages of our pipeline.

Next, we deconvolve the cluster frequency fields with the corresponding *Planck* isotropic instrumental beam at each field frequency (all of them available in the PLA) and with the HEALPix pixel transfer function. We perform these operations as simple multiplications in harmonic space, which, for the small fields we consider reduces to Fourier space.

2.2.2 Foreground suppression: constrained internal linear combination

Since we want to investigate how the primary CMB is deflected by each cluster, we need to suppress all the significant contributions to the observed signal that are on top of the primary, lensed CMB. We also need to reduce the instrument noise as much as possible. The most important foregrounds around a cluster are the thermal and the kinetic SZ effects. They add noise and, more troublingly, bias to the mass measurements (e.g., Melin and Bartlett 2015). The tSZ effect has the larger amplitude, in many cases larger than the cluster lensing signal by about an order of magnitude (Lewis and Challinor, 2006; Yoo et al., 2010). However, it has a very characteristic frequency dependence, different from the thermal CMB spectrum, which is independent of the cluster temperature in the non-relativistic limit (Rephaeli, 1995a). In this limit, the SZ spectrum is the same for all galaxy clusters. Thus, multi-frequency CMB observations can be combined in order to estimate and subtract the tSZ signal if the cluster gas is, to a good approximation, non-relativistic. As an illustration, the tSZ signal by the tSZ-brightest cluster in our sample, a cluster at $z = 0.0557$ with a tSZ signal-to-noise $q = 48.99$, can be seen by eye in Figure 2.1, which shows this cluster's cluster frequency fields. In addition, Figure 2.2 shows the cluster frequency fields at 143 GHz and 353 GHz ensemble averaged over our cluster sample, with the tSZ decrement (at 143 GHz) and increment (at 353 GHz) clearly visible.

On the other hand, the kinetic SZ signal, a Doppler-shift-induced contribution proportional to the cluster velocity along the line of sight, has the same spectral dependence as the primary CMB, so it cannot be removed with multi-frequency observations. Its amplitude can be of the same order of magnitude as the cluster lensing signal (Seljak and Zaldarriaga, 2000) and, for asymmetric clusters, may be expected to bias the mass measurement at some level. However, it has been argued through simulations that the bias is not significant for an experiment with the resolution and sensitivity of *Planck* and for clusters with realistic peculiar velocities (Melin and Bartlett, 2015). The kSZ does, however, add additional noise to the mass measurements.

We treat the tSZ effect as a signal to be explicitly removed from our fields and the instrumental noise and all the other foregrounds as noise on top of the primary, lensed CMB signal. In order to suppress the tSZ signal and to reduce the noise of our maps, we perform a constrained internal linear combination (CILC) with the six beam- and pixel-deconvolved cluster frequency fields, as described in Remazeilles et al. (2011). A CILC is a linear combination of a set of different frequency maps in which the weights are chosen in order to extract a signal with a known spectrum, minimising

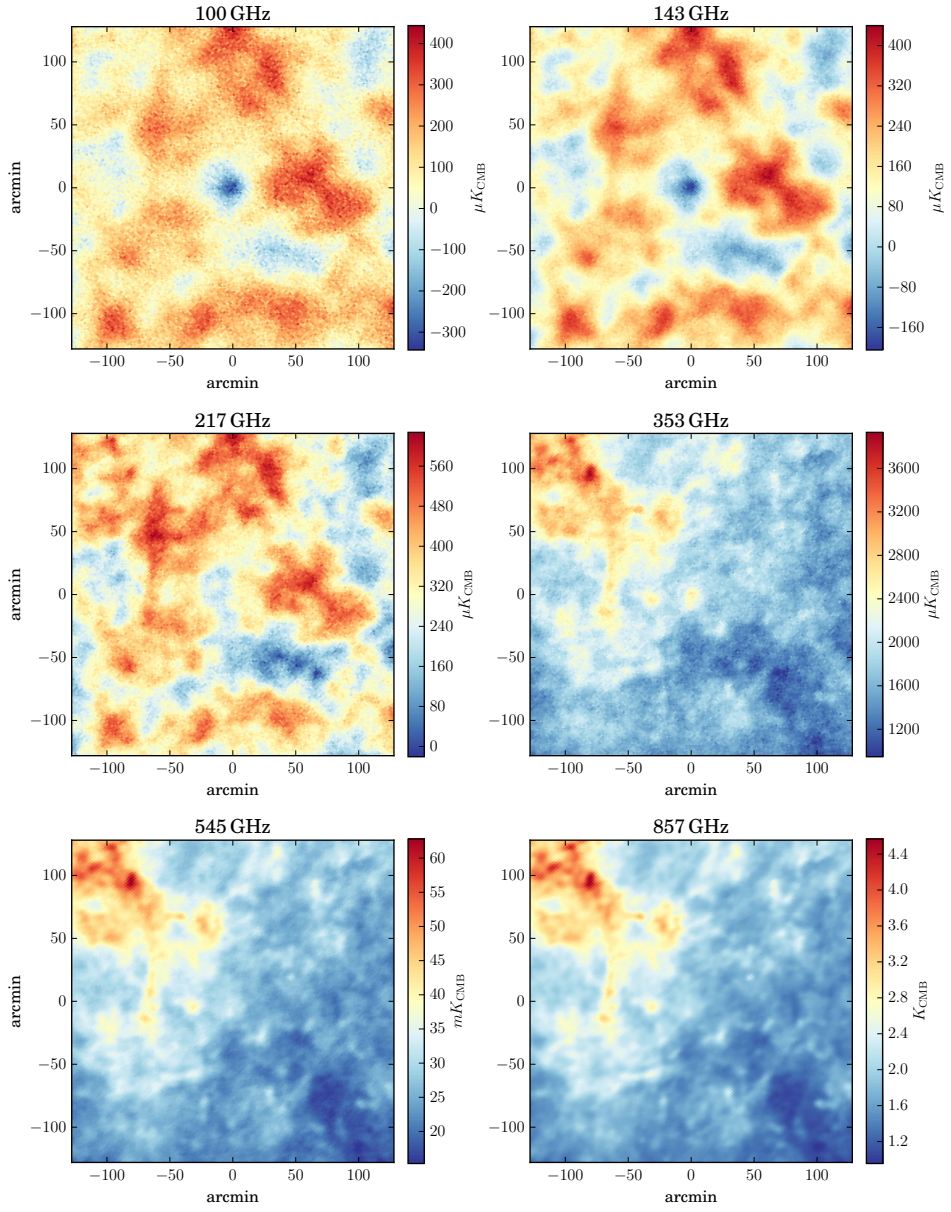


Fig. 2.1 Cluster frequency fields of our reference cluster. Note the tSZ decrement (increment) at frequencies below (above) 217 GHz, and the absence of a signal at this crossover frequency.

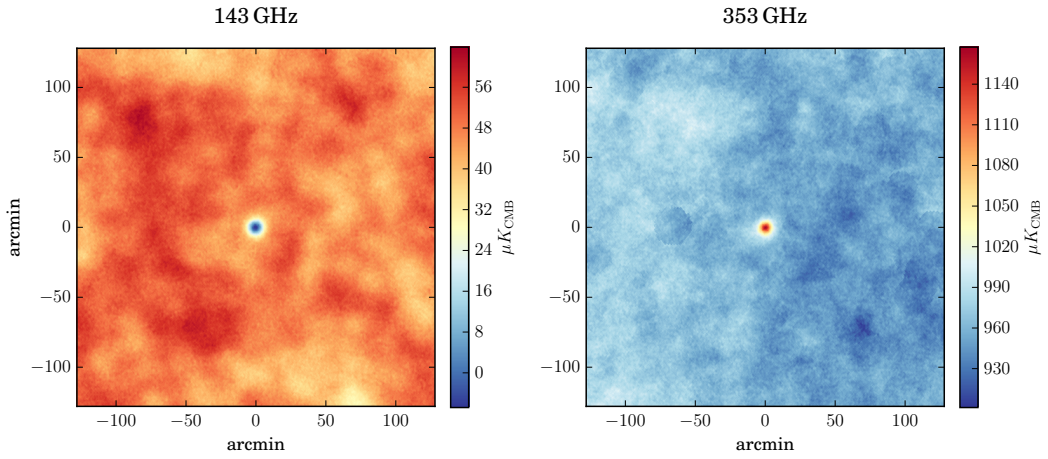


Fig. 2.2 Cluster frequency fields at 143 GHz and 353 GHz ensemble averaged over our cluster samples. The tSZ decrement is clearly visible at 143 GHz, and so is the increment at 353 GHz.

the variance of the output map with the constraint that another signal with a known spectral signature is completely removed. In this case, the signal to be extracted is the primary, lensed CMB, the signal to be removed is the tSZ signal, and the output is a single, tSZ-free temperature map around each cluster. We shall refer to these maps as 'cluster temperature fields'.

In more detail, in a CILC the observations (in our case, the six cluster frequency fields around a given cluster) are modelled as (Remazeilles et al., 2011)

$$\boldsymbol{x}(p) = \boldsymbol{a}s(p) + \boldsymbol{b}y(p) + \boldsymbol{n}(p). \tag{2.1}$$

Here, $\boldsymbol{x}(p)$ is a six-dimensional vector containing the values of the six frequency fields at the pixel p; $s(p)$ is the CMB map to be extracted, which also includes the kSZ signal as it has the same spectral dependence as the CMB; $y(p)$ is the tSZ contribution, to be removed; and $\boldsymbol{n}(p)$ denotes all the other unmodelled signals plus the instrumental noise. The fixed vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ describe the frequency dependence of the CMB (and kSZ) and tSZ signals, respectively, as integrated across the frequency bands of the experiment.

In units of the CMB temperature, the CMB vector $\boldsymbol{a} = (1, 1, 1, 1, 1, 1)^T$. In the non-relativistic limit, the tSZ vector $\boldsymbol{b}$ is the same for all clusters and has components for the *Planck* frequency bands that are given in Planck 2015 results VIII (2016).

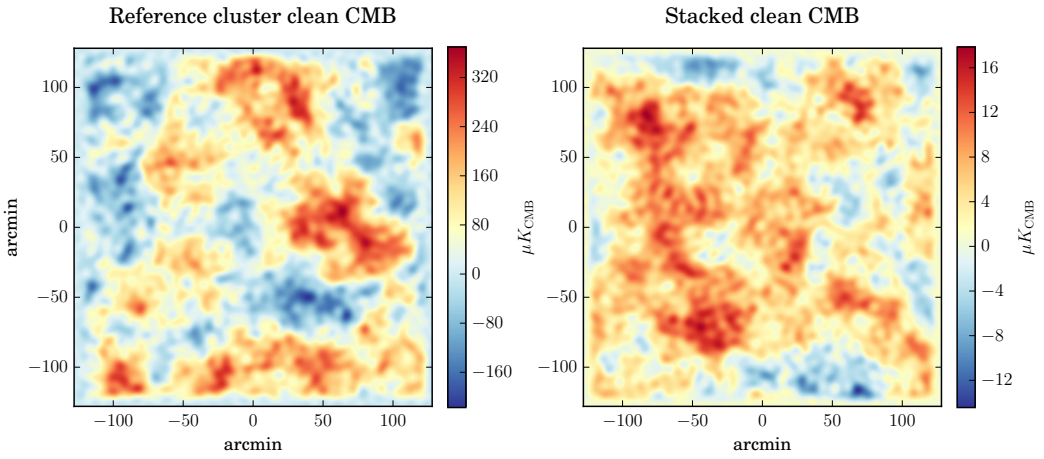


Fig. 2.3 Cluster temperature fields for our reference cluster (left panel), and averaged, or 'stacked', over our cluster sample (right panel). Both maps have been convolved with a beam with a FWHM of 2 arcmin for better visualisation.

Once $\boldsymbol{a}$ and $\boldsymbol{b}$ are known, a weight vector $\boldsymbol{w}$ can be constructed such that $\hat{s}(p) = \boldsymbol{w}^T\boldsymbol{x}(p)$ has unit response to the signal $s(p)$ and zero response to the contaminant $y(p)$. These impose the constraints $\boldsymbol{w}^T\boldsymbol{a} = 1$ and $\boldsymbol{w}^T\boldsymbol{b} = 0$. Further demanding that the effective noise variance be minimised leads to weights (Remazeilles et al., 2011)

$$\boldsymbol{w}^T = \frac{\left(\boldsymbol{b}^T\mathbf{R}^{-1}\boldsymbol{b}\right)\boldsymbol{a}^T\mathbf{R}^{-1} - \left(\boldsymbol{a}^T\mathbf{R}^{-1}\boldsymbol{b}\right)\boldsymbol{b}^T\mathbf{R}^{-1}}{\left(\boldsymbol{a}^T\mathbf{R}^{-1}\boldsymbol{a}\right)\left(\boldsymbol{b}^T\mathbf{R}^{-1}\boldsymbol{b}\right) - \left(\boldsymbol{a}^T\mathbf{R}^{-1}\boldsymbol{b}\right)^2}, \tag{2.2}$$

where $\mathbf{R} = \langle \boldsymbol{n}\boldsymbol{n}^T \rangle$ is the inter-frequency covariance matrix of the six frequency maps. Here, we assume this is constant across each cluster frequency field and estimate it empirically from the data by summing over pixels.[1]

The cluster temperature field of our reference cluster is shown in Figure 2.3 along with the sample average of our cluster temperature fields over our cluster sample. Both maps have been convolved with a Gaussian beam with a FWHM of 2 arcmin to aid visualisation. No tSZ residual signal is visible by eye in any of them.

[1]Note that adding (symmetric) outer products to $\mathbf{R}$ involving only $\boldsymbol{a}$ and/or $\boldsymbol{b}$ does not change the weights.

2.3 Lensing convergence estimation: quadratic estimators

The next step is to use our set of cluster temperature fields to estimate the lensing convergence around the clusters in our sample. We perform this estimation with a modified version of the well-known CMB lensing TT quadratic estimator, first described in Hu (2001), since for an experiment of the resolution and sensitivity of *Planck*, the TT quadratic estimator is very close to the optimal lensing estimator (Raghunathan et al., 2017).

CMB lensing by a fixed convergence $\kappa(\boldsymbol{x})$ destroys the statistical isotropy of the CMB, coupling Fourier modes with different wavevectors $\boldsymbol{l}$ in a κ-dependent way. The lensing quadratic estimators take advantage of this fact in order to estimate κ. The input of a quadratic estimator is a pair of observed CMB fields defined in the same region of the sky, which can consist of any combination of T, Q, or U; the input pair of the TT estimator is simply two copies of the same temperature map. Its output is an estimate $\hat{\kappa}$ of the convergence, which is unbiased at linear order in κ.

We use the following form of the TT quadratic estimator in order to estimate κ around a given cluster:

$$\hat{\kappa}(\boldsymbol{L}) = -N(\boldsymbol{L})\frac{2}{L^2}\int \frac{d^2\boldsymbol{x}}{2\pi} e^{-i\boldsymbol{L}\cdot\boldsymbol{x}}\boldsymbol{\nabla}\cdot\left[F_1(\boldsymbol{x})\nabla F_2(\boldsymbol{x})\right], \tag{2.3}$$

where $N(\boldsymbol{L})$ is a normalisation that can be written as

$$N(\boldsymbol{L}) = \frac{L^4}{4}\left(\int \frac{d^2\boldsymbol{l}}{(2\pi)^2}\frac{\left[(\boldsymbol{L}-\boldsymbol{l})\cdot\boldsymbol{L}C^{\tilde{T}\tilde{T}}_{|\boldsymbol{l}-\boldsymbol{L}|} + \boldsymbol{l}\cdot\boldsymbol{L}C^{\tilde{T}\tilde{T}}_{l}\right]^2}{2C^{\hat{T}\hat{T}}_{l}C^{\hat{T}\hat{T}}_{|\boldsymbol{l}-\boldsymbol{L}|}}\right)^{-1}. \tag{2.4}$$

Here, $C^{\tilde{T}\tilde{T}}_l$ is the power spectrum of the lensed CMB temperature, and $C^{\hat{T}\hat{T}}_l$ is the (total) power spectrum of the corresponding cluster temperature field. We have also defined two filtered temperature fields,

$$F_i(\boldsymbol{l}) \equiv f_i(l)\hat{T}(\boldsymbol{l}), \tag{2.5}$$

for $i = 1, 2$, where $\hat{T}(\boldsymbol{l})$ is the corresponding cluster temperature field and the filters are defined as

$$f_1(l) \equiv \frac{1}{C^{\hat{T}\hat{T}}_l}, \tag{2.6}$$

and

$$f_2(l) \equiv \begin{cases} \frac{C_l^{\tilde{T}\tilde{T}}}{C_l^{\hat{T}\hat{T}}} & \text{if } l \leq l_f, \\ 0 & \text{if } l > l_f. \end{cases} \tag{2.7}$$

Note that $F_1(\boldsymbol{l})$ is the inverse-variance-filtered temperature field, $F_2(\boldsymbol{l})$ is the Wiener-filtered temperature field (for $l \leq l_f$), and the quadratic estimator (2.3) involves the product of the former with the gradient of the latter.

Our TT quadratic estimator differs from the version originally described in Hu (2001) in two respects. First, we choose $f_2(l)$ to be zero for $l > l_f$. We take $l_f = 2000$. This ensures that the estimator remains unbiased in regions around galaxy clusters, where second and higher orders of κ become important and the usual filtering yields biased convergence estimates (Hu et al., 2007). In addition, we use the lensed CMB power spectrum, $C_l^{\tilde{T}\tilde{T}}$, in $f_2(l)$ and in the normalisation, rather than the unlensed CMB power spectrum, as this gives (approximately) the correct response of the estimator $\hat{\kappa}(\boldsymbol{L})$ to lenses at wavevector $\boldsymbol{L}$ averaged over all other lensing modes (Lewis et al., 2011; Hanson et al., 2011).

We estimate $C_l^{\hat{T}\hat{T}}$ empirically for each of our cluster temperature fields to account for the variation of residual foregrounds and instrumental noise across the sky. We calculate $C_l^{\tilde{T}\tilde{T}}$ with CAMB[2] assuming our fiducial cosmology. We then extract the central 256×256 pixels from our cluster temperature fields, and it is on these 128 arcmin $\times$ 128 arcmin fields ('reduced cluster temperature fields') that we reconstruct the lensing convergence. The reason for doing this is as follows. We first produce larger cluster temperature fields in order to obtain less noisy local power spectrum estimates. We then choose smaller fields on which to perform the lensing reconstruction to reduce the effect of neighbouring clusters and large-scale structure that is correlated with the cluster, which could contaminate the mass measurement of the central cluster.

Before performing the lensing reconstruction, we cosine apodise each reduced cluster temperature field over a width of 6.4 arcmin and then apply the same point source mask used to mask the corresponding set of frequency maps. We inpaint the source-masked regions with constrained Gaussian realisations, as detailed in Benoit-Lévy et al. (2013), so that the response of the quadratic estimator is the same as if there was no point-source masking. We account for the apodisation of the edges of the reduced cluster temperature fields by multiplying the reconstructed lensing convergence by a

[2]http://camb.info/

factor $1/\sqrt{f_4}$, where f_4 is given by (Benoit-Lévy et al., 2013)

$$f_4 = \frac{1}{N_{\rm pix}} \sum_i w_i^4. \tag{2.8}$$

Here, $N_{\rm pix}$ is the number of pixels in our reduced cluster temperature fields, w_i is the value of the apodisation mask at pixel i, and the sum is over all the pixels. Despite this normalisation being appropriate for measuring the power spectrum of statistically-isotropic lensing fields, we verified with simulations that we accurately recover the cluster convergence with this choice.

A mean field has to be subtracted from the lensing reconstruction to account for non-lensing sources of statistical anisotropy, which the quadratic estimator mistakes for lensing thereby biasing the reconstruction (e.g., Benoit-Lévy et al. 2013; Namikawa et al. 2013).

The sources of the mean field include inhomogeneous noise across the sky, anisotropic instrumental beams, and masks. All of these sources are present, to a certain degree, in our fields: *Planck* noise levels vary across the sky as does the residual foreground power in the CILC; the *Planck* (effective) instrumental beams are not perfectly isotropic; and our apodisation mask breaks isotropy. (The apodisation actually suppresses a larger mean field that would be present if there was no apodisation at all, due to the non-periodicity of the CMB cluster fields.) In order to subtract the mean field, we define the following modified quadratic estimator:

$$\hat{\kappa}_{\rm mod} = \frac{1}{\sqrt{f_4}} \left(\hat{\kappa} - \hat{\kappa}_{\rm mf} \right), \tag{2.9}$$

where, $\hat{\kappa}$ is the quadratic estimator presented above and $\hat{\kappa}_{\rm mf}$ is an estimate of the mean field.

The mean fields will differ across the cluster fields due to variation in the noise, effective beams and residual foreground levels. (We note that the largest contributions to the mean fields are from the masking, but this is approximately the same for all cluster fields, differing only in the point-source masking/inpainting.) We deal with this by estimating local mean fields from the data using reduced cluster temperature fields at random locations on sub-divisions of the sky. In detail, we divide the sky into 20 regions in bands of ecliptic latitude, since the instrument noise and beam contributions to the mean field are expected to vary significantly with latitude due to the *Planck* scan strategy. We identify each of the 10 bands falling in the northern-ecliptic hemisphere with the corresponding band located at its mirror location in the southern-ecliptic

hemisphere, yielding a total of 10 distinct bands. We then estimate a mean field within each band and subtract this from the raw convergence for clusters located with the band. To estimate the mean fields, we randomly select 16 384 fields with the same size and pixelisation as the cluster frequency fields, but centred on randomly-located points within each band and within the 2015 *Planck* 60 % Galactic mask, which was the mask used in the construction of the MMF3 cosmology catalogue. We apply the pipeline described above to the six *Planck* frequency maps for each randomly-centred frequency field and average the resulting estimated convergences (estimated using $\hat{\kappa}$) within each band. We tested with simulated CMB realisations that the number of random fields used to estimate the mean fields is large enough to allow us to ignore statistical errors in the mean fields in the mass estimates.

The estimated κ map around each cluster ('cluster reconstructed potential field') is therefore obtained by applying the quadratic estimator $\hat{\kappa}$ to the corresponding apodised and inpainted reduced cluster temperature field, subtracting an estimate of the mean field, and finally renormalising by $1/\sqrt{f_4}$. At fixed cluster convergence, the variance of this modified estimator is approximately given by

$$\langle \Delta\hat{\kappa}_{\rm mod}\left(\boldsymbol{L}\right) \Delta\hat{\kappa}^{*}_{\rm mod}\left(\boldsymbol{L}'\right)\rangle = \delta^{(2)}\left(\boldsymbol{L}-\boldsymbol{L}'\right) N_{\kappa}(\boldsymbol{L}), \tag{2.10}$$

where

$$N_{\kappa}(\boldsymbol{L}) = C_L^{\kappa\kappa} + N^{(0)}(L) + N^{(1)}(L). \tag{2.11}$$

Here, $C_L^{\kappa\kappa}$ is the lensing potential power spectrum (which we obtain from CAMB) and describes the variance from large-scale structure that is uncorrelated with the cluster; $N^{(0)}(L)$, known as the $N^{(0)}$ bias, describes Gaussian fluctuations of the CMB and noise that mimic the effects of lensing; and $N^{(1)}(L)$, known as the $N^{(1)}$ bias, depends linearly on $C_L^{\kappa\kappa}$ and arises from alternative couplings of the lensed CMB 4-point function with the two copies of the lensing quadratic estimator (Kesden et al., 2003). The $N^{(0)}$ bias is approximately equal to the estimator normalisation $N(\boldsymbol{L})$ given in Eq. (2.4), and we calculate the $N^{(1)}$ bias in the flat-sky limit following Kesden et al. (2003).

We perform the quadratic lensing reconstruction with the help of `quicklens`, a freely-available Python CMB lensing package.[3]

Figure 2.4 illustrates our convergence reconstruction process. The upper-left panel shows the ensemble-averaged, or stacked, cluster reconstructed convergence before mean field subtraction; the mean field is clearly visible by eye. The upper-right panel shows the average mean field, obtained by ensemble averaging over our 10 empirically-

[3]https://github.com/dhanson/quicklens

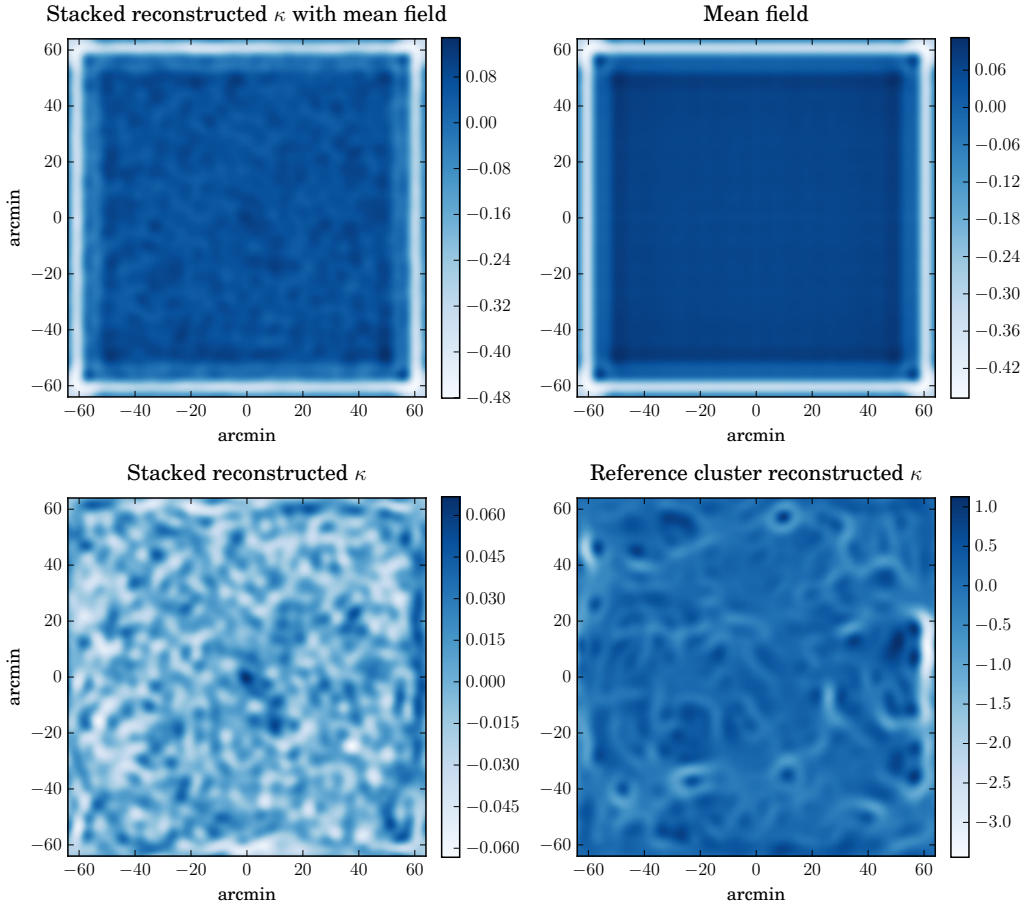


Fig. 2.4 Upper-left panel: ensemble-averaged, or stacked, cluster reconstructed convergence before mean field subtraction. Upper-right panel: average mean field, obtained by ensemble averaging over our 10 empirically-estimated mean fields. Lower-left panel: stacked cluster reconstructed convergence after mean field subtraction and renormalisation by $\sqrt{f_4}$. Lower-right panel: cluster reconstructed convergence field of our reference cluster. All four maps have been filtered with $N_\kappa(\mathbf{L})$ in Fourier space for better visualisation.

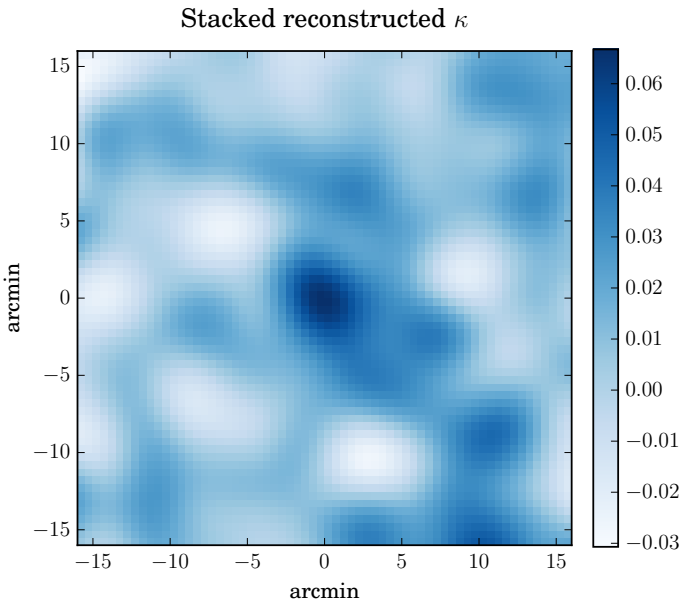


Fig. 2.5 Cut-out of the central region of the cluster reconstructed convergence fields ensemble averaged over our sample, filtered with $N_\kappa(\mathbf{L})$ in Fourier space for better visualisation.

estimated mean fields. The lower-left panel shows the stacked cluster reconstructed convergence after mean field subtraction and renormalisation by $\sqrt{f_4}$. A small high-convergence region can be clearly seen in the centre. For comparison, the lower-right panel shows the cluster reconstructed convergence field of our reference cluster, which illustrates how small the CMB lensing signal-to-noise per cluster is. All four maps have been filtered with $N_\kappa(\mathbf{L})$ in Fourier space for better visualisation.

In addition, a cut-out of the central region of the lower-left panel of Figure 2.4, i.e., of the stacked CMB lensing signal of the clusters in our sample, is shown in Figure 2.5. With a clear signal in the centre, this map can be seen as a visual summary of our detection of CMB lensing.

2.4 Matched filtering

We estimate a cluster mass from each cluster reconstructed convergence field using a matched filter, as presented in Melin and Bartlett (2015) and used in the *Planck* analysis in Planck 2015 results XXIV (2016). The method consists of matched filtering the estimated convergence, $\hat{\kappa}_{\rm mod}$, with a model for the cluster convergence (the lensing

potential is matched filtered in Melin and Bartlett 2015, but both procedures are equivalent).

2.4.1 Cluster model

Following Planck 2015 results XXIV (2016), we choose for the cluster mass model a truncated NFW profile (Navarro et al., 1997):

$$\rho(r) = \begin{cases} \frac{\rho_0}{(r/r_s)(1+r/r_s)^2} & \text{if } r \leq R_{\text{trunc}}, \\ 0 & \text{if } r > R_{\text{trunc}}, \end{cases} \tag{2.12}$$

where ρ_0 is a characteristic density; r_s is a characteristic scale radius; and R_{trunc} is the truncation radius.

In order to be consistent with Planck 2015 results XXIV (2016), we define the cluster mass in terms of M_{500}, the mass contained with a radius R_{500} such that the mean enclosed mass is 500 times that of the critical density at the cluster's redshift, $\rho_c(z)$. Imposing the definition of M_{500},

$$\rho_0 = \rho_c(z)\frac{500}{3}\frac{c_{500}^3}{\ln(1+c_{500}) - c_{500}/(1+c_{500})}, \tag{2.13}$$

where the concentration parameter is defined by $c_{500} = R_{500}/r_s$. As in Melin and Bartlett (2015), we assume that $c_{500} = 3$, although it is known that it actually varies weakly with cluster mass and redshift (Muñoz-Cuartas et al., 2011; Diemer and Kravtsov, 2015). We further follow Melin and Bartlett (2015) in their choice of truncation radius: $R_{\text{trunc}} = 5R_{500}$. Fixing c_{500}, for our purposes a cluster is completely specified by two parameters, e.g., M_{500} and redshift z.

The lensing convergence of the cluster at sky position $\boldsymbol{\theta}$ from its centre, $\kappa_{\text{cl}}(\boldsymbol{\theta})$, is related to the (physical) surface density at projected position $\boldsymbol{r}$ from the centre, $\Sigma_{\text{cl}}(\boldsymbol{r})$, through

$$\kappa_{\text{cl}}(\boldsymbol{\theta}) = \Sigma_{\text{cl}}(\boldsymbol{r})/\Sigma_{\text{crit}}(z), \tag{2.14}$$

where the critical surface density at redshift z is

$$\Sigma_{\text{crit}}(z) = \frac{c^2}{4\pi G}\frac{d_{A,\text{CMB}}}{d_{A,\text{cl}}d_{A,\text{cl-CMB}}}, \tag{2.15}$$

which involves the physical angular diameter distances to the CMB last-scattering surface, to the cluster, and between the cluster and the CMB. The projected position

at the cluster, $\boldsymbol{r}$, is related to the angular position, $\boldsymbol{\theta}$, through $\boldsymbol{r} = d_A(z)\boldsymbol{\theta} = d_{A,\mathrm{cl}}\boldsymbol{\theta}$. The projected surface density can be written as

$$\Sigma_{\mathrm{cl}}(\boldsymbol{r}) = 2\rho_0 r_s \int_{r/r_s}^{5c_{500}} \frac{dx}{(1+x)^2\sqrt{x^2-(r/r_s)^2}}, \tag{2.16}$$

for $|\boldsymbol{r}| \leq R_{\mathrm{trunc}}$. For a fixed concentration, the integral only depends on $\theta/\theta_s = |\boldsymbol{r}|/r_s$, where the angular scale radius is $\theta_s = r_s/d_A(z)$. It follows that the cluster convergence can be written in the form

$$\kappa_{\mathrm{cl}}(\boldsymbol{\theta}) = \kappa_0 \kappa_{\mathrm{t}}(\boldsymbol{\theta};\theta_s), \tag{2.17}$$

where the circularly-symmetric template function $\kappa_{\mathrm{t}}(\boldsymbol{\theta};\theta_s)$ depends only on θ/θ_s. We choose the normalisation such that $\kappa_{\mathrm{t}} = 1$ at the scale radius, so that κ_0 is the cluster convergence there. It follows that

$$\kappa_0 = \frac{2\rho_0 r_s}{\Sigma_{\mathrm{crit}}(z)} \int_1^{5c_{500}} \frac{dx}{(1+x)^2\sqrt{x^2-1}}, \tag{2.18}$$

and so, at fixed concentration,

$$\kappa_0 \theta_s^2 \propto \frac{M_{500}}{\Sigma_{\mathrm{crit}}(z) d_A^2(z)}, \tag{2.19}$$

i.e., the integrated convergence within an aperture of radius θ_s is proportional to the cluster mass.

2.4.2 Matched filter

Given an estimate of a cluster angular size θ_s, we filter its reconstructed convergence field with the template κ_{t} to form an estimator for κ_0:

$$\hat{\kappa}_0 = \left[\int \frac{d^2\boldsymbol{L}}{2\pi} \frac{|\kappa_{\mathrm{t}}(\boldsymbol{L})|^2}{N_\kappa(\boldsymbol{L})}\right]^{-1} \int \frac{d^2\boldsymbol{L}}{2\pi} \frac{\hat{\kappa}(\boldsymbol{L})\kappa_{\mathrm{t}}^*(\boldsymbol{L})}{N_\kappa(\boldsymbol{L})}, \tag{2.20}$$

where $N_\kappa(\boldsymbol{L})$ is the variance of $\hat{\kappa}(\boldsymbol{L})$ given in Eq. (2.11). The inverse-variance weighting of $\hat{\kappa}$ ensures this estimator has minimum variance, which can be calculated to be

$$\sigma_{\kappa_0}^2 \equiv \left\langle (\hat{\kappa}_0 - \kappa_0)^2 \right\rangle = \frac{1}{2\pi}\left[\int \frac{d^2\boldsymbol{L}}{2\pi} \frac{|\kappa_{\mathrm{t}}(\boldsymbol{L})|^2}{N_\kappa(\boldsymbol{L})}\right]^{-1}. \tag{2.21}$$

In order to use the filter, an estimate of the cluster angular scale is needed. We use the SZ mass proxy $M_{\rm SZ}$ of each cluster, as provided in the MMF3 catalogue, as the filter mass, $M^{\rm fid}$, from which we derive an SZ angular size $\theta_s^{\rm fid}$:

$$M^{\rm fid} = 500\frac{4\pi}{3}\left[d_A(z)c_{500}\theta_s^{\rm fid}\right]^3\rho_c(z). \tag{2.22}$$

Since $\kappa_0\theta_s^2 \propto M_{500}$, we can write an estimator of the cluster's mass as

$$\hat{M}_{500} = M^{\rm fid}\left(\frac{\hat{\kappa}_0}{\kappa_0^{\rm fid}}\right), \tag{2.23}$$

where $\kappa_0^{\rm fid}$ is the template's convergence at its scale radius $\theta_s^{\rm fid}$. The standard deviation of this estimator is simply

$$\sigma_{M_{500}} = M^{\rm fid}\left(\frac{\sigma_{\kappa_0}}{\kappa_0^{\rm fid}}\right). \tag{2.24}$$

If the filter matches exactly the cluster's true profile, $\hat{M}_{500}$ is an unbiased estimator of the cluster's true mass, M_{500}. However, as we now discuss, $\hat{M}_{500}$ is biased at linear order in any mismatch between the template and the true profile. We can avoid such linear bias by instead working with the signal-to-noise on κ_0 (or, equivalently, on M_{500}), defining the observable

$$p_{\rm obs} \equiv \hat{\kappa}_0/\sigma_{\kappa_0} = \hat{M}_{500}/\sigma_{M_{500}}. \tag{2.25}$$

2.4.3 Biases from template errors

We consider three possible deviations of the template $\kappa_{\rm t}$ from the true cluster convergence, which will bias the estimator to some extent: mismatch between the filter fiducial angular scale and the true cluster angular scale, actual mismatch between the filter profile and the true profile, and misplacing of the filter position with respect to the cluster's true position (which we refer to as miscentering).

Let us first consider the effect of mismatch between the filter fiducial angular scale and the true cluster angular scale. Consider a cluster of true size θ_s, but filtered with $\theta_s^{\rm fid}$. The expected value of the signal-to-noise is

$$\langle p_{\rm obs}\rangle = \sqrt{2\pi}\kappa_0\left[\int\frac{d^2\boldsymbol{L}}{2\pi}\frac{|\kappa_{\rm t}(\boldsymbol{L};\theta_s^{\rm fid})|^2}{N_\kappa(\boldsymbol{L})}\right]^{-1/2}\int\frac{d^2\boldsymbol{L}}{2\pi}\frac{\kappa_{\rm t}(\boldsymbol{L};\theta_s)\kappa_{\rm t}^*(\boldsymbol{L};\theta_s^{\rm fid})}{N_\kappa(\boldsymbol{L})}, \tag{2.26}$$

where κ_0 is the true value of the convergence at the true scale radius. Recalling that the template convergence $\kappa_{\rm t}(\boldsymbol{\theta};\theta_s)$ depends only on θ/θ_s, we can write

$$\kappa_{\rm t}(\boldsymbol{\theta};\theta_s) = f(\theta/\theta_s), \tag{2.27}$$

where the function f is circularly symmetric. The Fourier transform $\kappa_{\rm t}(\boldsymbol{L};\theta_s)$ is therefore related to the 2D Fourier transform of f through

$$\kappa_{\rm t}(\boldsymbol{L};\theta_s) = \theta_s^2 f(\theta_s \boldsymbol{L}). \tag{2.28}$$

Substituting into Eq. (2.26), we have

$$\langle p_{\rm obs}\rangle = \sqrt{2\pi}\kappa_0\theta_s^2 \left[\int \frac{d^2\boldsymbol{L}}{2\pi}\frac{f^2(\theta_s^{\rm fid}\boldsymbol{L})}{N_\kappa(\boldsymbol{L})}\right]^{-1/2} \int \frac{d^2\boldsymbol{L}}{2\pi}\frac{f(\theta_s\boldsymbol{L})f(\theta_s^{\rm fid}\boldsymbol{L})}{N_\kappa(\boldsymbol{L})}, \tag{2.29}$$

The dependence of $\langle p_{\rm obs}\rangle$ on $\theta_s^{\rm fid}$ can be assessed by writing $\theta_s^{\rm fid} = \theta_s + \delta\theta_s$ and series expanding in $\delta\theta_s$. For lens reconstruction with Planck data, the reconstruction noise is large on typical cluster scales, i.e., for $\theta_s|\boldsymbol{L}| \gtrsim 1$, so $\delta\theta_s|\boldsymbol{L}| \ll 1$ for the modes that dominate the integrals in Eq. (2.29) and the series expansion is expected to converge rapidly. It is straightforward to show that there is no linear dependence of $\langle p_{\rm obs}\rangle$ on $\delta\theta_s$, so the bias from using the incorrect cluster angular size in the filtering is only second order in the size error.

We can check numerically that this bias on $p_{\rm obs}$ from adopting an incorrect filter size is indeed small. In particular, we consider a cluster with a mass $M_{500} = 0.5\times10^{15}M_\odot$ at redshift $z = 0.2$ and with our truncated NFW profile as its true convergence profile. We compute $\langle p_{\rm obs}\rangle$ with Eq. (2.29), taking the template to be the true convergence profile but scaled at a number of different angular scales $\theta_s^{\rm fid}$. For the variance $N_\kappa(\boldsymbol{L})$ we use that for a TT quadratic estimator reconstruction for an idealised *Planck*-like experiment with a Gaussian beam of full-width at half-maximum equal to 5 arcmin and noise level of 45 μK arcmin. The cluster masses associated with each of the filter angular scales considered, $M^{\rm fid}$, computed with Eq. (2.22), and their corresponding $\langle p_{\rm obs}\rangle$ values are shown in Table 2.1. It can be seen that value of the filter angular scale $\theta_s^{\rm fid}$, parametrised in Table 2.1 in terms of the filter mass proxy $M^{\rm fid}$, has little impact on $\langle p_{\rm obs}\rangle$ if it is within a reasonable range of the cluster's true angular scale θ_s. Since our filter mass proxies for each cluster are their SZ mass proxies, as provided in the MMF3 catalogue, and given that we expect them to be about $1-b \approx 0.7$ times the cluster true masses, we conclude that the choice of filter angular scale has a negligible impact on our analysis. This would have not been the case if we had chosen

Table 2.1 Results of our numerical test of the dependence of the expected value of the matched filter output on the filter angular scale $\theta_s^{\rm fid}$. The cluster considered here has a mass $M_{500} = 0.5 \times 10^{15} M_\odot$, is at redshift $z = 0.2$, and has our truncated NFW profile as its convergence. The templates with which we filter its convergence have functionally the same profile and are also placed at redshift $z = 0.2$, but have a number of different masses $M^{\rm fid}$, which, at fixed redshift, correspond uniquely to a set of different filter angular scales $\theta_s^{\rm fid}$. The first column shows the filter masses, $M^{\rm fid}$, that we have considered; the second column shows the corresponding values of $\langle p_{\rm obs} \rangle$; and the third column, the corresponding values of the expected value of the filter mass estimate, $\langle \hat{M}_{500} \rangle$. It can be seen that $\langle p_{\rm obs} \rangle$ is very insensitive to $M^{\rm fid}$ (and, therefore, to $\theta_s^{\rm fid}$) if $M^{\rm fid}$ is reasonably close to the cluster's true mass, whereas $\langle \hat{M}_{500} \rangle$ has a significant dependence on $M^{\rm fid}$.

$M^{\rm fid}$ $10^{15} M_\odot$	$\langle p_{\rm obs} \rangle$	$\langle \hat{M}_{500} \rangle$ $10^{15} M_\odot$
0.01	0.2023	0.25
0.10	0.2136	0.36
0.35	0.2164	0.46
0.50	0.2165	0.50
1.00	0.2159	0.58
10.00	0.2055	0.98

to use the filter mass estimates directly, as was done in the CMB lensing calibration in Planck 2015 results XXIV (2016). Indeed, the third column of Table 2.1 shows the expected values of the cluster mass estimates at each filter angular scale, $\langle \hat{M}_{500} \rangle$. It can be observed that a significant bias in the mass estimates appears if the filter angular scale $\theta_s^{\rm fid}$ is different from the truth, θ_s – in our test, that corresponding to $M_{500} = 0.5 \times 10^{15} M_\odot$. Indeed, $\hat{M}_{500}$ is biased at linear order, while $p_{\rm obs}$ is biased only at second order. We therefore use $p_{\rm obs}$ as our lensing mass observable.

Let us now consider the more general case of mismatch between the filter profile and the mean true cluster profile at given cluster mass and redshift. Following an analogous argument to that for the bias from mismatch in angular size, we argue that profile mismatch biases $\hat{M}_{500}$ to linear order, but $p_{\rm obs}$ only to second order. If the mismatch is small, the bias on $p_{\rm obs}$ will therefore also be small.

It should also be noted that, even if the mean cluster profile at given mass M_{500} and redshift z matches our model, the profile of a given real cluster with mass M_{500} and redshift z can be thought of as a 'noisy' realisation of such a mean profile, with the scatter arising from triaxiality and large-scale structure correlated with the cluster. This intrinsic scatter is not modelled in our matched filtering process, which only

includes variance arising from the lensing reconstruction noise and from uncorrelated large-scale structure, so is still present in our $p_{\rm obs}$ measurements. As we explain in Section 3.4, we do, however, account for intrinsic scatter in the mass measurements in our likelihood.

Finally, let us consider the possibility of miscentering of the filter with respect to the cluster's true position. We centre the filter at the midpoint of each reduced cluster temperature field, corresponding to the SZ-estimated position of the cluster centre as provided in the *Planck* MMF3 catalogue. These positions are, however, not known perfectly. The *Planck* MMF3 catalogue includes an estimate of the 95 % confidence interval of the positional uncertainty of each cluster due to the *Planck* beam. Its mean and median values over the cluster sample are both around 2.4 arcmin, which implies a typical offset of about 1 arcmin. In addition, there is another source of miscentering arising from the offset between the SZ-estimated centre and the centre of mass of each cluster, which is where we ought to place our template. This offset is difficult to estimate, but we expect it to be smaller than the offset due to the beam. Indeed, using hydrodynamical simulations, Gupta et al. (2017) find that for about 80 % of the clusters the typical value of this offset is of about $0.04R_{500}$. For a typical cluster in our sample, this translates into an offset of about 0.25 arcmin, which, added in quadrature to the offset due to the beam, does not increase it significantly. We can therefore expect a typical miscentering angle, d, of about 1 arcmin. Assuming that the cluster's mean true profile exactly matches the filter, the filter and the true mean profile are related by

$$\kappa_{\rm t}^{\rm true}(\boldsymbol{L}) = \kappa_{\rm t}(\boldsymbol{L};\theta_s)e^{i\boldsymbol{L}\cdot\boldsymbol{d}}, \tag{2.30}$$

where $\boldsymbol{d}$ is the centering error. It is straightforward to show that the bias this induces in $p_{\rm obs}$ is only second order in d (see Appendix A for details). We expect the leading-order bias to be negative, following the $O(\boldsymbol{L}\cdot\boldsymbol{d})^2$ term in the expansion of $\exp(i\boldsymbol{L}\cdot\boldsymbol{d})$, with a value of around 3 % for $d = 1$ arcmin in the case in which the filter variance $N_\kappa(\boldsymbol{L})$ is appropriate for a TT quadratic estimator reconstruction for a *Planck*-like experiment. In this scenario, the next-order non-vanishing contribution to the bias is $O(d^4)$ and is at the 10^{-4} level, and thus negligible.

We do not explicitly correct for these small, second-order residual biases in our lensing measurements. We do, however, include an effective CMB lensing mass bias parameter in our likelihood analysis, as explained in Section 3.4, which we marginalise over to account for the biases.

2.5 Conclusion

In this chapter we have measured the CMB lensing signal-to-noise ratios of the 433 clusters of the *Planck* CMB lensing calibration sample using the six highest frequency (100–857 GHz) Full-Mission 2015 *Planck* temperature maps. To make these measurements, first, for each cluster we have produced a clean temperature map around its sky location, explicitly projecting out its tSZ signal with a constrained internal linear combination. Then, we have obtained lensing convergence estimates using a modified version of the TT quadratic estimator in which a cut-off at $l = 2000$ in the gradient leg is imposed, which ensures that the estimator remains unbiased in the central regions of clusters. Care has been taken to subtract the mean field due to inhomogeneous noise across the sky, anisotropic instrumental beams, and masks, which we have estimated empirically from random cut-outs. Finally, we have matched filtered our reconstructed convergence fields with a truncated NFW model, obtaining a CMB lensing signal-to-noise measurement for each cluster. We have argued that the signal-to-noise is a better-behaved mass observable than the mass estimate, in our case virtually insensitive to the matched filter input angular scale.

These CMB lensing signal-to-noise measurements, which constitute an early detection of CMB lensing by clusters, are then used in the cosmological analysis of *Planck* galaxy clusters that we present in Chapter 3, in which they play the crucial role of calibrating the absolute cluster mass scale.

Chapter 3

Cosmological constraints from *Planck* galaxy clusters with CMB lensing mass bias calibration

3.1 Introduction

As discussed in Chapter 1, galaxy clusters, the largest gravitationally-bound structures in the Universe, are powerful cosmological probes. In particular, their abundance as a function of mass and redshift is very sensitive to the mean matter density of the Universe, which can be parametrised by $\Omega_{\rm m}$, and to the amplitude of the matter perturbations, which can be characterised by σ_8, the root mean square of the linear density fluctuations smoothed on a scale of $8\,h^{-1}\,$Mpc. In recent years, large Sunyaev–Zel'dovich (SZ) cluster surveys, in which clusters are detected through their thermal-SZ (tSZ) signature (Sunyaev and Zeldovich, 1972), have provided useful observations of this abundance from which cosmological constraints have been obtained (e.g., Staniszewski et al. 2009; Mantz et al. 2010; Hasselfield et al. 2013; Bleem et al. 2015; Planck 2015 results XXVII 2016; de Haan et al. 2016; Bocquet et al. 2019; Salvati et al. 2018). SZ surveys are particularly interesting because the change in surface brightness of the cosmic microwave background (CMB) due to the SZ effect does not decrease with cluster redshift, allowing for the detection of high-redshift galaxy clusters given enough resolution. In addition, the observational selection of the cluster sample is typically straightforward to model, simplifying the extraction of cosmological information.

A crucial element in analysing the abundance of galaxy clusters ('cluster counts' analysis) is the observational determination of cluster masses. In SZ surveys, the

mass of a galaxy cluster is typically estimated directly from CMB data through its mass-dependent SZ signal: the SZ signal is said to be a proxy of the galaxy cluster mass. However, the scaling relations between a cluster's mass and its SZ signal are not very well determined, and usually they need to be calibrated for each survey. This determination of cluster masses currently provides the largest source of uncertainty when obtaining cosmological information from galaxy clusters (Planck 2015 results XXIV, 2016; Pratt et al., 2019).

The *Planck* experiment detected about 1200 galaxy clusters via their SZ signature (Planck 2015 results XXVII, 2016). A subsample of 439 such clusters, known as the MMF3 cosmology sample, was used in a counts analysis in order to constrain, amongst other parameters and models, $\Omega_{\rm m}$ and σ_8 within the context of a standard spatially-flat ΛCDM cosmology (Planck 2015 results XXIV, 2016). This analysis proceeded as follows. Each cluster was characterised by two observables: SZ signal-to-noise ratio, q, and redshift, z. The signal-to-noise ratio q, a proxy of the cluster mass, was measured from *Planck* data for each cluster. On the other hand, the redshifts of most clusters in the sample (433) were measured by follow-up observations. These two cluster observables were then binned on a grid in the q–z plane. The number of clusters within each cell was modelled as being independently Poisson distributed in a likelihood analysis, with the mean number of clusters being the theoretically-predicted quantities dependent on cosmology.

Theory predicts the number of clusters as a function of their redshift and true mass (we use M_{500}, the mass within a radius R_{500}, inside which the mean matter density is 500 times that of the critical density at the cluster's redshift, $\rho_{\rm c}(z)$). To connect these predictions to the expected counts in the q–z plane, the analysis of Planck 2015 results XXIV (2016) used the SZ-mass scaling relations from Arnaud et al. (2010), with parameters calibrated with X-ray mass estimates of a subsample of clusters of the MMF3 cosmology sample (see Planck 2013 results XXIX 2014 and Planck 2015 results XXVII 2016 for the calibration details). However, it is known that the X-ray mass estimates are typically biased low. To account for this, a mass bias factor was introduced, such that in the X-ray-calibrated scaling relations used in the likelihood, the X-ray-derived masses were replaced by the scaled true masses, $(1-b)M_{500}$, where $1-b$ parametrises the mass bias (a somewhat awkward parametrisation introduced in Planck 2013 results XX 2014). There are several possible sources of this bias: it is known that X-ray cluster mass estimates are typically biased low at a significant level due to their being obtained under the assumption of hydrostatic equilibrium within the cluster, an assumption that can be violated in several scenarios (Nagai

et al., 2007a; Piffaretti and Valdarnini, 2008; Meneghetti et al., 2010); there can also be observational systematic effects in the X-ray mass estimates and selection effects biasing X-ray-selected samples relative to SZ-selected samples (Planck 2015 results XXIV, 2016).

To obtain cosmological constraints, it is necessary to calibrate the mass bias parameter $1-b$ since this determines the overall cluster mass scale. For example, an increase in σ_8, which increases the number of clusters above a given true mass, can be offset by a reduction in $1-b$, which makes the true masses of the observed clusters larger, preserving the expected number of observed clusters. In the analysis of Planck 2015 results XXIV (2016), the mass bias was estimated from three independent calibrations. In each of these calibrations, lensing mass estimates of the clusters of a subsample of the MMF3 cosmology sample were directly compared to the corresponding masses that are obtained from the X-ray calibrated SZ-mass scaling relations if no mass bias is assumed. Lensing probes the total cluster mass distribution without relying on any of the assumptions that underlie X-ray mass estimates, and so is arguably the most robust way to measure cluster masses to date. Two of the calibrations used in Planck 2015 results XXIV (2016) were from galaxy weak-lensing mass estimates of a small number of clusters in the MMF3 cosmology sample: $1-b=0.688\pm0.072$ from the Weighing the Giants (WtG, von der Linden et al. 2014) programme; and $1-b=0.78\pm0.07$ from the Canadian Cluster Comparison Project (CCCP, Hoekstra et al. 2015). Both galaxy weak-lensing calibrations support a significant mass bias. The third calibration used CMB lensing mass estimates of most of the clusters in the MMF3 cosmology sample (433) and reported $1/(1-b)=0.99\pm0.19$ (Planck 2015 results XXIV, 2016), which is consistent with their being no mass bias. These three different calibrations were then used as priors on the mass bias parameter in the cluster counts likelihood analysis. They affect the results significantly, particularly the constraints on σ_8. As shown in Figure 7 of Planck 2015 results XXIV (2016), the absence of a mass bias, a scenario suggested by the CMB lensing calibration, leads to parameter constraints in the $\Omega_{\rm m}$–σ_8 plane that are in tension with the corresponding constraints derived from the *Planck* CMB power spectra in the ΛCDM model. On the other hand, the two galaxy weak-lensing mass bias calibrations alleviate the tension significantly. Model-dependent constraints on the mass bias can be obtained by combining cluster counts with the CMB power spectra. Within ΛCDM, the combination of the *Planck* temperature and polarisation power spectra (the *Planck* 2015 TT,TE,EE+lowP likelihood) and the *Planck* cluster-counts likelihood gives a posterior distribution $1-b=0.58\pm0.04$ (Planck 2015 results XXIV, 2016). This confirms that a significant mass bias is required if the ΛCDM cosmology

favoured by the primary CMB anisotropies is to be consistent with the observed counts of galaxy clusters.

The mass calibration from joint analysis of the cluster counts and primary CMB power spectra was recently updated in Planck 2018 results VI (2018) to use the *Planck* 2018 TT,TE,EE+lowE+lensing likelihood (the cluster-counts likelihood was unchanged from Planck 2015 results XXIV 2016). The main relevant change is the improved constraints on large-scale polarisation due to the use of the data from the Planck High Frequency Instrument rather than the less precise data from the Low Frequency Instrument used in the 2015 likelihoods. This change favours a lower central value for the optical depth to reionization, τ, with the new constraint also being around twice as precise. Lowering τ reduces the fluctuation amplitude to preserve the amplitude of the CMB spectra on smaller scales and leads to a lower predicted value of σ_8. Consistency with the observed cluster counts then requires a larger value of $1-b$ (i.e., less mass bias), with Planck 2018 results VI (2018) reporting $1-b=0.62\pm0.03$.

In this chapter we revisit the *Planck* cluster counts analysis for the particular case of the CMB lensing mass bias calibration. After measuring again the masses of all the clusters in the *Planck* CMB lensing calibration sample through their CMB lensing signature, we argue that a mass bias calibration like the one presented in Planck 2015 results XXIV (2016) is biased by several effects.

In order to account for these effects, we present an alternative approach in which the CMB lensing mass estimates of each cluster, along with q and z, are directly incorporated into a likelihood that is able to constrain jointly $\Omega_{\rm m}$, σ_8, and $1-b$ at once. We present the constraints obtained from our analysis within the ΛCDM model, finding good agreement with the constraints from the *Planck* CMB power spectra.

This chapter is organised as follows. First, in Section 3.2 we detail the cluster sample used in our analysis. Next, in Section 3.3 we explain why a mass bias calibration such as the one presented in Planck 2015 results XXIV (2016) is, in general, biased, and in Section 3.4 we present a likelihood that, by combining the CMB lensing mass estimates of each cluster in our sample with the corresponding q and z, has the power to constrain $\Omega_{\rm m}$, σ_8, and $1-b$ jointly in an unbiased way. We validate this approach with simulated data in Section 3.5. Our parameter results are presented and discussed in Section 3.6. We conclude in Section 3.7. A series of appendices provide additional details of the origin of the selection bias that can affect the mass calibration (Appendix B), propose a method for dealing with uncertainty in cluster centering in future analyses (Appendix A), and discuss the relation of our likelihood to a simple Poisson counts likelihood in the space of cluster observables (Appendix C).

3.2 Data

For our cosmological analysis we use the clusters of the *Planck* MMF3 cosmology sample. This sample contains a total of 439 clusters and is the sample that was used in the baseline analysis in Planck 2015 results XXIV (2016). Details of the construction of this sample are given in Planck 2015 results XXVII (2016) and Planck 2013 results XXIX (2014). Briefly, clusters are detected through their tSZ signature across the six highest frequency *Planck* channels, and selected by imposing that their tSZ signal-to-noise ratio $q > 6$, and that they are in the area of the sky left unmasked in the analysis (65 % of the sky). The cluster catalogue is publicly available in the *Planck* Legacy Archive (hereafter, PLA)[1].

As a measure of the SZ signal for each cluster in the sample, we follow Planck 2015 results XXIV (2016) and use the signal-to-noise ratio, q. This was measured by the *Planck* Collaboration for all the clusters of the MMF3 cosmology sample through their MMF3 pipeline (Planck 2015 results XXVII, 2016). We similarly use the cluster redshifts obtained by the Planck Collaboration from ancillary data and follow-up observations, as described in Planck 2015 results XXVII (2016).

In order to calibrate the mass bias, we also incorporate in our analysis CMB lensing mass estimates of all the clusters in the MMF3 cosmology sample with measured redshift. This subsample consists of a total of 433 clusters, and it is the same sample as the one used in the CMB lensing calibration presented in Planck 2015 results XXIV (2016). In the following, we will refer to it as the *Planck* CMB lensing calibration sample. We directly estimate the masses of all the clusters in this sample from *Planck* 2015 full-mission temperature maps, as detailed in Chapter 2. For those clusters with redshift information, the *Planck* Collaboration also provide estimates of each cluster's mass, $M_{\rm SZ}$. These are obtained by combining measurements of the cluster SZ observables θ_{500}, the cluster angular scale, and Y_{500}, the SZ flux within a θ_{500} aperture, which are generally strongly correlated given the resolution of *Planck*, with X-ray-calibrated fiducial scaling relations to break the degeneracy; see Planck 2015 results XXVII (2016) for further details. From this conditional estimate of Y_{500}, the mass proxy $M_{\rm SZ}$ can be derived from the same scaling relations. As emphasised in Planck 2015 results XXVII (2016), $M_{\rm SZ}$ should be regarded as the expected *hydrostatic mass* of a cluster, given the assumed scaling relations, cluster redshift and distribution of Y_{500} and θ_{500} derived from the data. In this work, we only use $M_{\rm SZ}$ to provide an initial angular scale for the matched filter used to estimate the cluster mass from CMB

[1] http://pla.esac.esa.int/pla/

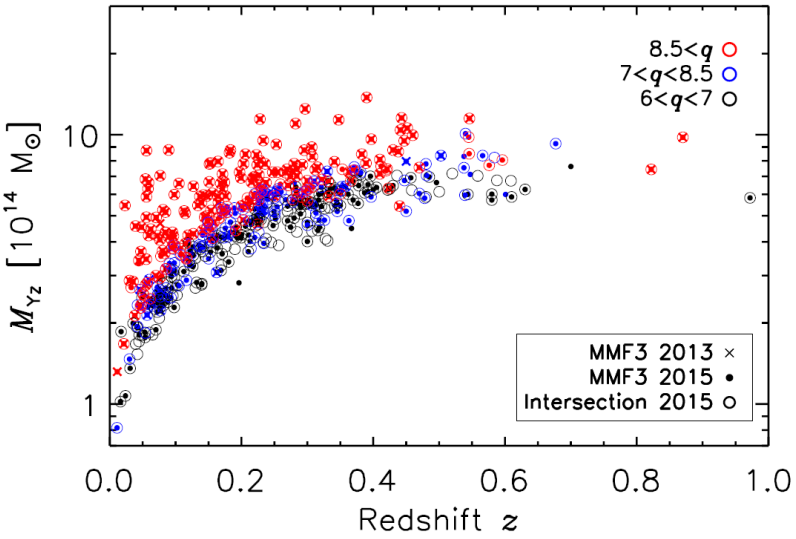


Fig. 3.1 Mass–redshift distribution of the 2015 MMF3 cosmology sample, of its 2013 version, and of the intersection sample (see Planck 2015 results XXIV 2016). Here, $M_{\rm Yz}$ is the *Planck* SZ mass proxy. Figure credit: Planck 2015 results XXIV (2016).

lensing data (Section 2.4), and in Section 3.3 where we compare with the lensing masses to provide a simple (but biased) estimate of the hydrostatic mass bias following Planck 2015 results XXIV (2016).

For illustration, the mass–redshift distribution of the 2015 MMF3 cosmology sample, used for the cosmological analysis presented in this chapter, is shown in Figure 3.1, along with other *Planck* cluster samples.

3.3 Pitfalls of simple estimation of the SZ mass bias

As described in Planck 2015 results XXIV (2016), the mass bias $1-b$ is introduced in the *Planck* cluster counts analysis in order to account for a possible bias in the X-ray-derived masses used to calibrate the SZ–mass scaling relations. It enters the analysis by multiplying the true cluster mass, M_{500}, wherever M_{500} appears in the scaling relations.

As proposed in Planck 2015 results XXIV (2016), a possible way to estimate the mass bias is to construct the following estimator of $1/(1-b)$:

$$\frac{\hat{1}}{1-b} = \frac{M_{\rm lens}}{M_{Y_z}}, \tag{3.1}$$

where $M_{\rm lens}$ is a lensing estimate of the cluster mass and M_{Y_z} is the corresponding scaling-relation-derived mass. Averaging over an ensemble of clusters, this estimator is an unbiased estimator of $1/(1-b)$ if the following conditions are met: (i) the $M_{\rm lens}$ measurements are unbiased estimates of M_{500}; (ii) the M_{Y_z} measurements are unbiased estimates of the mean SZ mass at a given M_{500}, which is what we can define as $(1-b)M_{500}$; and (iii) the scatter on the mean SZ mass at a given M_{500} can be neglected, that is, the M_{Y_z} measurements are precisely $(1-b)M_{500}$. Indeed, if these conditions are satisfied,

$$\left\langle \frac{\hat{1}}{1-b} \right\rangle = \left\langle \frac{M_{\rm lens}}{M_{Y_z}} \right\rangle = \frac{\langle M_{\rm lens} \rangle}{M_{Y_z}} = \frac{M_{500}}{(1-b)M_{500}} = \frac{1}{1-b}, \tag{3.2}$$

where angular brackets denote averaging over the cluster sample with some appropriate weighting to minimise the variance. It should be noted that even with such weighting, this estimator is not in general the optimal estimator of the mass bias.

The conditions needed for this estimator of $1/(1-b)$ to be unbiased are not fully met for the *Planck* CMB lensing calibration sample, yielding, in general, an incorrect estimate of the mass bias. A detailed quantification of the size of the error, and the extent to which the underlying assumptions are invalid, is beyond the scope of this work. However, here we offer a brief, qualitative description of the unsuitability of two of the three assumptions that are necessary for the estimator to be accurate.

First, as already noted in Chapter 2, it is known that, even in the limit of an unbiased lensing convergence reconstruction around a cluster, the matched filtering process generally gives biased mass estimates due to, e.g., mismatch between the true and template convergence profiles and miscentering. If unaccounted for, as in the calibration presented in Planck 2015 results XXIV (2016), these biases in the lensing mass estimates propagate directly into $1/(1-b)$.

Secondly, at a given M_{500}, the SZ mass estimates in the sample, M_{Y_z}, are biased high compared to the true mean SZ mass, $(1-b)M_{500}$. This is due to clusters in our calibration sample being selected through a cutoff on the SZ signal-to-noise, q (plus the additional neglect of six clusters for which there are no redshift measurements available, although this should have very little impact on the modelling of the selection of our sample). This selection effect, which has already been studied in the literature (e.g., Mantz et al. 2010; Allen et al. 2011; Nagarajan et al. 2019), can be understood as follows. Since M_{Y_z} are noisy realisations of the mean SZ mass at a given M_{500}, close to the selection cutoff the mean of the SZ masses that get included in the sample for a given M_{500} is necessarily larger than the true population mean $(1-b)M_{500}$. As a

consequence, M_{Y_z} becomes a biased estimator of $(1-b)M_{500}$ close to the cutoff for our sample, even if before selection it was unbiased. We therefore expect the simple estimator (3.1) to underestimate $1/(1-b)$, i.e., $1-b$ to be biased high.

This selection effect also has some impact on the lensing mass estimates, despite these not intervening in the selection of the sample. This is because the intrinsic scatter in lensing masses is expected to be correlated with that in the SZ masses to a certain extent. Indeed, for a given cluster, both the SZ and the lensing mass estimates are obtained from quantities integrated along the line of sight. Thus, it is reasonable to expect some correlation in their intrinsic scatter, e.g., due to cluster triaxiality. Correlations between SZ and galaxy weak-lensing masses have indeed been observed in realistic simulations of galaxy clusters, with a reported intrinsic correlation coefficient as high as 0.8 (Shirasaki et al., 2016). Similarly, we can expect the scatter in the CMB lensing mass of a cluster to correlate with that in its SZ mass, biasing the mean lensing mass in the sample high if the correlation is positive. Nevertheless, we expect this to be a small effect for current CMB lensing mass estimates since the measurement errors dominate over the intrinsic scatter.

More details of the selection effect discussed above can be found in Appendix B.

3.4 Joint likelihood analysis of cluster SZ and mass measurements

In this section we develop our Bayesian model to constrain cosmological parameters, notably $\Omega_{\rm m}$ and σ_8, and the SZ mass bias (along with several other nuisance parameters) from joint analysis of the SZ data and cluster mass estimates from CMB lensing. The central object is the likelihood, $\mathscr{L}$, giving the probability of the data given the model parameters. The data we use are as follows: (i) the total number of clusters in the MMF3 cosmology sample, N; (ii) the cluster locations on the celestial sphere, $\{\hat{\boldsymbol{n}}_i\}$; and (iii) the cluster redshifts, $\{z_{{\rm obs},i}\}$, as measured by the *Planck* collaboration in follow-up observations, SZ signal-to-noise ratios, $\{q_{{\rm obs},i}\}$, as measured by the *Planck* collaboration through their MMF3 pipeline, and CMB lensing signal-to-noise ratios, $\{p_{{\rm obs},i}\}$, as measured in this work (see Chapter 2). We collectively refer to $z_{{\rm obs},i}$, $q_{{\rm obs},i}$, and $p_{{\rm obs},i}$ as the 'mass data point' D_i of cluster i. Since redshift measurements are not available for six of the clusters in the MMF3 cosmology sample, the mass data point of each such cluster reduces to $q_{{\rm obs},i}$.

The likelihood can be written as the probability density function of N, $\underline{\hat{\boldsymbol{n}}}$, and $\underline{D}$, $P\left(N, \underline{\hat{\boldsymbol{n}}}, \underline{D}\right)$, where $\underline{\hat{\boldsymbol{n}}}$ is the vector whose ith-component is $\hat{\boldsymbol{n}}_i$, and $\underline{D}$ is the vector

whose i-component is D_i. This probability density function can be decomposed as

$$P\left(N, \underline{\hat{\boldsymbol{n}}}, \underline{D}\right) = P(\underline{D}|N, \underline{\hat{\boldsymbol{n}}}) P(\underline{\hat{\boldsymbol{n}}}|N) P(N). \tag{3.3}$$

The first term, $P(\underline{D}|N, \underline{\hat{\boldsymbol{n}}})$, which we shall refer to as $\mathscr{L}_1$, is the probability of obtaining the cluster mass data points $\underline{D}$ given that N clusters have been included in our sample (the MMF3 cosmology sample) and that their sky locations have been found to be $\underline{\hat{\boldsymbol{n}}}$. The dependence on sky location is important, since foregrounds and the *Planck* instrumental noise vary significantly across the sky. The second term, $P(\underline{\hat{\boldsymbol{n}}}|N)$, which we shall denote with $\mathscr{L}_2$, is the probability that, given N clusters have been included in our sample, they are located at the sky locations $\underline{\hat{\boldsymbol{n}}}$. Finally, the third term, $P(N)$, which we shall refer to as $\mathscr{L}_3$, is the probability of including a total of N clusters in our cluster sample.

Our likelihood is a natural way to extend the SZ counts formalism in order to incorporate the CMB lensing mass measurements to allow for self-calibration of the SZ mass bias. In the rest of this section, we develop the three factors in Eq. (3.3), making clear the parameters on which they depend. In Appendix C we show how, in general, a likelihood like ours is equivalent to a Poisson counts likelihood in z–$q_{\rm obs}$–$p_{\rm obs}$ space (in a suitable limit) and how, in particular, our likelihood can be reduced by marginalisation over $p_{\rm obs}$ to a likelihood similar to the *Planck* SZ counts likelihood.

3.4.1 $\mathscr{L}_1$: Mass data likelihood

In order to construct $\mathscr{L}_1 = P(\underline{D}|N, \underline{\hat{\boldsymbol{n}}})$, we assume that each cluster in our sample is statistically independent of the others. This assumption was also made in Planck 2015 results XXIV (2016), where it was claimed that the impact on the results of the correlations between the different clusters in the sample due to large-scale clustering is weak. As a consequence, we can write $P(\underline{D}|N, \underline{\hat{\boldsymbol{n}}})$ as a product of the probability density functions of the mass data point of each cluster in the sample:

$$P(\underline{D}|N, \underline{\hat{\boldsymbol{n}}}) = \prod_{i=1}^{N} P(D_i|\mathrm{in}, \hat{\boldsymbol{n}}_i). \tag{3.4}$$

Recall that D_i is the mass data point of each cluster, $D_i = (q_{{\rm obs},i}, p_{{\rm obs},i}, z_{{\rm obs},i})$ for the clusters with known redshift, and $D_i = (q_{{\rm obs},i})$ for the clusters with unknown redshift. The condition in $P(\underline{D}|N, \underline{\hat{\boldsymbol{n}}})$ that N clusters are included in the sample is translated into each $P(D_i|\mathrm{in}, \hat{\boldsymbol{n}}_i)$ as the condition that each of the clusters is included in the sample. We denote this condition with "in" in Eq. (3.4) and throughout.

We take $P(D_i|\mathrm{in}, \hat{\boldsymbol{n}}_i)$ to have the same functional form for all the clusters with known redshift. The same applies for the clusters with unknown redshift. In this case, $P(D_i|\mathrm{in}, \hat{\boldsymbol{n}}_i)$ is obtained by marginalising the probability density function of the mass data point of a cluster with known redshift over the corresponding $p_{\mathrm{obs},i}$ and $z_{\mathrm{obs},i}$.

In the following we describe how we construct $P(D_i|\mathrm{in}, \hat{\boldsymbol{n}}_i)$ for a cluster with known redshift.[2] To avoid clutter in the notation, hereafter we will drop the cluster index, i. First, we note that

$$P(D|\mathrm{in}, \hat{\boldsymbol{n}}) = \frac{P(\mathrm{in}|D, \hat{\boldsymbol{n}})P(D|\hat{\boldsymbol{n}})}{P(\mathrm{in}|\hat{\boldsymbol{n}})}. \tag{3.5}$$

Given the selection criterion applied in the construction of the MMF3 cosmology sample, $q_{\mathrm{obs}} > 6$, $P(\mathrm{in}|D, \hat{\boldsymbol{n}})$ is simply a step function at $q_{\mathrm{obs}} = 6$. In order to obtain the other two terms in Eq. (3.5), we adopt a hierarchical model to link each cluster mass data point, $D = (q_{\mathrm{obs}}, p_{\mathrm{obs}}, z_{\mathrm{obs}})$, to the true cluster mass, M_{500}, and redshift, z, and then assume a probability distribution for M_{500} and z (which is what theory can predict).

Our hierarchical model has two layers between M_{500} and z and the mass data point, D. First, q_{obs}, p_{obs}, and z_{obs} are thought of as noisy realisations of their 'true' values, q_{t}, p_{t}, and z_{t}, respectively. These are the values of these quantities that would be obtained after averaging over all sources of 'observational noise'. We refer to this layer as 'observational scatter', and specify what we mean exactly by observational noise below. Second, q_{t}, p_{t}, and z_{t} are understood as noisy realisations of some mean $\bar{q}\,(M_{500}, z, \hat{\boldsymbol{n}})$, $\bar{p}\,(M_{500}, z, \hat{\boldsymbol{n}})$, and $\bar{z}\,(M_{500}, z, \hat{\boldsymbol{n}})$, respectively, which are specified at given M_{500}, z, and cluster sky location, $\hat{\boldsymbol{n}}$. We refer to this layer as 'intrinsic scatter', and specify its physical sources below.[3] With this hierarchical model in mind, we can write the probability density function followed by the mass data point of one single cluster as

$$\begin{aligned} P(D|\hat{\boldsymbol{n}}) &= P(q_{\mathrm{obs}}, p_{\mathrm{obs}}, z_{\mathrm{obs}}|\hat{\boldsymbol{n}}) \\ &= \int dq_{\mathrm{t}} dp_{\mathrm{t}} dz_{\mathrm{t}} dM_{500} dz \; [P(q_{\mathrm{obs}}, p_{\mathrm{obs}}, z_{\mathrm{obs}}|q_{\mathrm{t}}, p_{\mathrm{t}}, z_{\mathrm{t}}, \hat{\boldsymbol{n}}) \\ &\qquad \times P(q_{\mathrm{t}}, p_{\mathrm{t}}, z_{\mathrm{t}}|M_{500}, z, \hat{\boldsymbol{n}}) P(M_{500}, z|\hat{\boldsymbol{n}})] \,, \end{aligned} \tag{3.6}$$

[2] For generality, we develop the likelihood formalism allowing for scatter in the redshift estimates, although in our implementation with the *Planck* clusters we can ignore such scatter for those clusters with redshift information.

[3] More precisely, as explained below, $\ln \bar{q}\,(M_{500}, z, \hat{\boldsymbol{n}})$ and $\ln \bar{p}\,(M_{500}, z, \hat{\boldsymbol{n}})$ are the means of $\ln q_{\mathrm{t}}$ and $\ln p_{\mathrm{t}}$, respectively.

where the first factor of the integrand accounts for the observational scatter, the second one accounts for the intrinsic scatter, and the last one is the unconditioned probability density function followed by M_{500} and z.

Since the selection criterion applied in the construction of the MMF3 cosmology sample depends exclusively on the value of $q_{\rm obs}$, and given the functional form of the intrinsic and observational scatters, which will be described below, the remaining term on the right of Eq. (3.5), $P(\mathrm{in}|\hat{\boldsymbol{n}})$, can be written using a simplified version of the hierarchical model. Indeed, we can write

$$P(\mathrm{in}|\hat{\boldsymbol{n}}) = \int dq_{\rm obs} dq_{\rm t} dM_{500} dz \; [P(\mathrm{in}|q_{\rm obs}, \hat{\boldsymbol{n}}) P(q_{\rm obs}|q_{\rm t}, \hat{\boldsymbol{n}}) \times P(q_{\rm t}|M_{500}, z, \hat{\boldsymbol{n}}) P(M_{500}, z|\hat{\boldsymbol{n}})] \,. \quad (3.7)$$

Here, as we will see, the probability distributions governing the intrinsic and observational scatters on the q variables, $P(q_{\rm t}|M_{500}, z, \hat{\boldsymbol{n}})$ and $P(q_{\rm obs}|q_{\rm t}, \hat{\boldsymbol{n}})$, respectively, can be readily obtained from those of the full hierarchical model by marginalising over the corresponding p and z variables. Also, as in Eq. (3.5), $P(\mathrm{in}|q_{\rm obs}, \hat{\boldsymbol{n}})$ is a step function with the step located at $q_{\rm obs} = 6$.

In the following we describe how we compute the mean signal-to-noise at given M_{500} and z for the SZ, CMB lensing, and redshift measurements, and the specific models we adopt for the intrinsic scatter, the observational scatter, and for the unconditioned probability density function of M_{500} and z.

Mean quantities: $\bar{q}$, $\bar{p}$, and $\bar{z}$

Let us consider a cluster with some given values of M_{500} and z, and at sky location $\hat{\boldsymbol{n}}$. As stated above, we assume that these three variables specify some mean values of the SZ signal-to-noise, the CMB lensing mass signal-to-noise, and the redshift of each cluster, $\bar{q}\,(M_{500}, z, \hat{\boldsymbol{n}})$, $\bar{p}\,(M_{500}, z, \hat{\boldsymbol{n}})$, and $\bar{z}\,(M_{500}, z, \hat{\boldsymbol{n}})$.

First, let us consider the mean CMB lensing mass signal-to-noise. As detailed in Chapter 2, for each cluster we filter the noisy reconstruction of the lensing convergence around its location with a truncated NFW profile in order to obtain an estimate of its lensing mass and the signal-to-noise on this.

In our hierarchical model, we assume that clusters of some M_{500} and z have, on averaging over cluster shape and correlated and uncorrelated large-scale structure, projected mass distributions and so lensing convergences described by the truncated NFW profile specified in Chapter 2 at the true redshift, z, but at a biased mass. Matched filtering the reconstructed convergence with an assumed profile to estimate

the mass can introduce further biases due to miscentering and profile mismatch. We assume that the net effect of all sources of bias is to give a mean lensing mass signal-to-noise for clusters of true mass M_{500}, redshift z, and sky location $\hat{\boldsymbol{n}}$ equal to that for a truncated NFW cluster at redshift z and mass $(1-b_{\rm CMBlens})M_{500}$ filtered at this same (biased) mass scale and redshift:

$$\bar{p}\left(M_{500}, z, \hat{\boldsymbol{n}}\right) = \frac{(1-b_{\rm CMBlens})M_{500}}{\sigma_{M_{500}}\left[(1-b_{\rm CMBlens})M_{500}, z, \hat{\boldsymbol{n}}\right]}. \tag{3.8}$$

Here, $\sigma_{M_{500}}\left[(1-b_{\rm CMBlens})M_{500}, z\right]$ is the matched filter noise, given by Eq. (2.24). This introduces a dependence of $\mathscr{L}_1$ on the sky location of each cluster, since the lensing reconstruction noise varies across the sky. We assume that the lensing mass bias $b_{\rm CMBlens}$ is constant across the sample. At the resolution and noise levels of *Planck*, $\bar{p}$ scales roughly linearly with $(1-b_{\rm CMBlens})M_{500}$.

For the mean SZ signal-to-noise, $\bar{q}$, since we use the measured SZ signal-to-noise ratios $q_{\rm obs}$ as provided in the *Planck* MMF3 catalogue, we compute $\bar{q}$ at given cluster mass, redshift, and sky location in exactly the same way as the *Planck* team (see Planck 2013 results XXIX 2014, Planck 2015 results XXIV 2016, and Planck 2015 results XXVII 2016). That is, we write the mean SZ signal-to-noise at M_{500}, z, and sky location $\hat{\boldsymbol{n}}$ as

$$\bar{q}(M_{500}, z, \hat{\boldsymbol{n}}) = \frac{\bar{Y}_{\rm SZ}\left[(1-b_{\rm SZ})M_{500}, z\right]}{\sigma_{\rm f}\left[\theta_{500}((1-b_{\rm SZ})M_{500}, z), \hat{\boldsymbol{n}}\right]}. \tag{3.9}$$

Here, $\sigma_{\rm f}(\theta_{500}, \hat{\boldsymbol{n}})$ is the noise of the multifrequency matched filter used to detect the clusters evaluated at the SZ angular scale of the cluster, θ_{500}, where

$$\theta_{500} = \theta_\star \left(\frac{h}{0.7}\right)^{-2/3} \left(\frac{(1-b_{\rm SZ})M_{500}}{3\times 10^{14} M_\odot}\right)^{1/3} E^{-2/3}(z) \left(\frac{d_A(z)}{500{\rm Mpc}}\right)^{-1}. \tag{3.10}$$

The constant $\theta_\star = 6.997\,{\rm arcmin}$, $E(z) = H(z)/H_0$, where $H(z)$ is the Hubble parameter and the Hubble constant $H_0 = 100h\,{\rm km\,s^{-1}\,Mpc^{-1}}$, and $d_A(z)$ is the angular diameter distance to redshift z. The SZ mass bias parameter $b_{\rm SZ}$ is introduced to account for any differences between the X-ray mass estimates that are used in the calibration of the SZ scaling relations and the true masses, as discussed in Section 3.1. The mean integrated Comptonisation parameter in Eq. (3.9), $\bar{Y}_{\rm SZ}$, is given by the scaling relation

(Planck 2015 results XXIV, 2016)

$$E^{-\beta}(z)\left(\frac{d_A^2(z)\bar{Y}_{\rm SZ}}{10^{-4}\,{\rm Mpc}^2}\right) = Y_\star\left(\frac{h}{0.7}\right)^{-2+\alpha}\left(\frac{(1-b_{\rm SZ})M_{500}}{6\times10^{14}M_\odot}\right)^{\alpha}, \tag{3.11}$$

where $Y_\star$, α, and β are parameters that need to be calibrated. We compute $\sigma_{\rm f}(\theta_{500},\hat{\boldsymbol{n}})$ by interpolating over the tabulated values used in the likelihood analyses in Planck 2015 results XXIV (2016); their likelihood code is freely available as part of the COSMOMC package.[4] As for the CMB lensing case, $\sigma_{\rm f}(\theta_{500},\hat{\boldsymbol{n}})$ introduces an additional dependence of $\mathscr{L}_1$ on the sky location of each cluster.

Finally, as in Planck 2015 results XXIV (2016), we take the mean redshift at M_{500}, z, and $\hat{\boldsymbol{n}}$ to be simply z, i.e.,

$$\bar{z}\,(M_{500},z,\hat{\boldsymbol{n}}) = z. \tag{3.12}$$

Intrinsic scatter

The first layer in our hierarchical model is what we refer to as intrinsic scatter. For the SZ and CMB lensing measurements, by intrinsic scatter we mean all sources of statistical scatter on the signal-to-noise measurements that are not incorporated in the noise budget of the matched filtering process. For the CMB lensing measurements, these include deviations of the cluster convergence profile from its assumed mean profile at M_{500} and z, deviations which often exhibit a triaxial nature, and contributions to the observed convergence profile from correlated large-scale structure. For the SZ measurements, intrinsic scatter includes deviations from the mean Comptonisation profile, also often of a triaxial nature. Finally, as in Planck 2015 results XXIV (2016), we assume that there is no intrinsic scatter on the redshift measurements.

In Planck 2015 results XXIV (2016) a log-normal model for the intrinsic scatter of the SZ signal-to-noise was adopted. Intrinsic scatter on lensing-derived cluster masses has also been shown to be approximately log-normal (see, e.g., Becker and Kravtsov 2011), and the SZ and CMB lensing measurements are also expected to be intrinsically correlated, since they are both obtained from quantities integrated along the line of sight (see, e.g., Shirasaki et al. 2016). We therefore adopt the following probability distribution for the intrinsic scatter:

$$P(q_{\rm t},p_{\rm t},z_{\rm t}|M_{500},z,\hat{\boldsymbol{n}}) = \frac{1}{q_{\rm t}p_{\rm t}}P(\ln q_{\rm t},\ln p_{\rm t}|M_{500},z,\hat{\boldsymbol{n}})\delta(z_{\rm t}-z), \tag{3.13}$$

[4] https://cosmologist.info/cosmomc/

where the delta function reflects the absence of intrinsic scatter on the redshift measurements, and where $P(\ln q_\mathrm{t}, \ln p_\mathrm{t}|M_{500}, z, \hat{\boldsymbol{n}})$ is a bivariate Gaussian with means $\ln \bar{q}(M_{500}, z, \hat{\boldsymbol{n}})$ and $\ln \bar{p}(M_{500}, z, \hat{\boldsymbol{n}})$, respectively, and a covariance matrix that we parametrise with its two associated standard deviations, σ_SZ and σ_CMBlens, respectively, and a correlation coefficient, $r_\mathrm{CMBlens-SZ}$. We take σ_SZ, σ_CMBlens, and $r_\mathrm{CMBlens-SZ}$ to be constants.

Observational scatter

The second layer in our hierarchical model is what we call observational scatter. For the CMB lensing and SZ measurements, it accounts for the scatter caused by what is treated as observational noise in the matched filtering processes that yield the values of q_obs and p_obs of each cluster. For example, for the CMB lensing masses, observational scatter arises from the reconstruction noise and large-scale structure that is uncorrelated with the cluster. We assume that the observational scatter is Gaussian for both SZ and lensing measurements. This assumption was also made for the SZ measurements in Planck 2015 results XXIV (2016), and should be also a good approximation for the CMB lensing measurements, since they are obtained by matched filtering a reconstructed convergence map, which involves a sum over many reconstructed modes. We also assume that there is no correlation between the observational scatter of the CMB lensing measurements and that of the SZ measurements. Finally, as in Planck 2015 results XXIV (2016), we assume that there is no observational scatter on z_t.

We can therefore write the probability distribution function accounting for the observational scatter as

$$P(q_\mathrm{obs}, p_\mathrm{obs}, z_\mathrm{obs}|q_\mathrm{t}, p_\mathrm{t}, z_\mathrm{t}, \hat{\boldsymbol{n}}) = P(q_\mathrm{obs}|q_\mathrm{t})P(p_\mathrm{obs}|p_\mathrm{t})\delta(z_\mathrm{obs} - z_\mathrm{t}), \tag{3.14}$$

where $P(q_\mathrm{obs}|q_\mathrm{t})$ and $P(p_\mathrm{obs}|p_\mathrm{t})$ are Gaussians with means q_t and p_t, respectively, and unit standard deviations. The delta function in Eq. (3.14) accounts for the absence of observational scatter on the redshift measurements.

Unconditioned distribution of M_{500} and z

The remaining factor of the integrand in Eq. (3.6) is $P(M_{500}, z|\hat{\boldsymbol{n}})$. This is the probability density function followed by M_{500} and z with no conditions imposed, other than the sky location $\hat{\boldsymbol{n}}$. It does not depend on $\hat{\boldsymbol{n}}$ and is simply proportional to the

halo mass function, $d^4N/(d^3V dM_{500})$, times the volume element, $d^3V/(dz d^2\Omega)$:

$$P(M_{500}, z|\hat{\boldsymbol{n}}) \propto \frac{d^4N}{d^3V dM_{500}}(M_{500}, z)\frac{d^3V}{dz d^2\Omega}(z). \tag{3.15}$$

As in Planck 2015 results XXIV (2016), we use the halo mass function from Tinker et al. (2008).

Probability of inclusion, or renormalisation

Finally, let us revisit the probability for a cluster to be included in the sample, $P(\text{in}|\hat{\boldsymbol{n}})$, as written down in Eq. (3.7). From Section 3.4.1, it should now be apparent that such a decomposition is possible, and that $P(q_{\rm obs}|q_{\rm t}, \hat{\boldsymbol{n}})$ is just $P(q_{\rm obs}|q_{\rm t})$ in Eq. (3.14), and $P(q_{\rm t}|M_{500}, z, \hat{\boldsymbol{n}})$ is just $P(q_{\rm t}, p_{\rm t}, z_{\rm t}|M_{500}, z, \hat{\boldsymbol{n}})$ in Eq. (3.13), marginalised over $p_{\rm t}$ and $z_{\rm t}$.

3.4.2 $\mathscr{L}_2$: Sky location likelihood

The second factor of our total likelihood, $\mathscr{L}_2$ (see Eq. 3.3), is $P(\underline{\hat{\boldsymbol{n}}}|N)$, the probability that, given N clusters have been included in our sample (the MMF3 cosmology sample), their sky locations are $\underline{\hat{\boldsymbol{n}}}$. Since we assume that clusters are statistically independent from each other, it is simply given by

$$P(\underline{\hat{\boldsymbol{n}}}|N) = N! \prod_{i=1}^{N} \frac{1}{\bar{N}} \frac{d^2\bar{N}}{d^2\Omega}(\hat{\boldsymbol{n}}_i), \tag{3.16}$$

where $d\bar{N}/d\Omega$ is the mean number of clusters in the sample per solid angle, and $\bar{N}$ is the mean total number of clusters in the sample, i.e.,

$$\bar{N} = \int d^2\hat{\boldsymbol{n}} \frac{d^2\bar{N}}{d^2\Omega}(\hat{\boldsymbol{n}}), \tag{3.17}$$

where the integration is performed across the region of the sky left unmasked in the construction of the MMF3 cosmology sample. We include the $N!$ factor in Eq. (3.16) to reflect the fact that the ordering of the elements in $\underline{\hat{\boldsymbol{n}}}$ does not matter.[5] The mean number of clusters in the sample per solid angle can be written as

$$\frac{d^2\bar{N}}{d^2\Omega}(\hat{\boldsymbol{n}}) = \int dq_{\rm obs} dM_{500} dz\, P\left(q_{\rm obs}|M_{500}, z, \hat{\boldsymbol{n}}\right) \frac{d^4N}{d^3V dM_{500}} \frac{d^3V}{dz d^2\Omega}. \tag{3.18}$$

[5] The integration measure over the data for N clusters is then $(1/N!)\prod_{i=1}^{N} dD_i d^2\hat{\boldsymbol{n}}_i$.

Here, $q_{\rm obs}$ is integrated from 6 to infinity (to meet the inclusion selection), $d^4N/(d^3V dM_{500})$ is the halo mass function, and $d^3V/(dz d^2\Omega)$ is the volume element (see Section 3.4.1). The probability density

$$P\left(q_{\rm obs}|M_{500}, z, \hat{\boldsymbol{n}}\right) = \int dq_{\rm t}\, P\left(q_{\rm obs}|q_{\rm t}\right) P\left(q_{\rm t}|M_{500}, z, \hat{\boldsymbol{n}}\right), \tag{3.19}$$

which involves the (marginalised) probability density functions for the observational and intrinsic scatter introduced in Section 3.4.1.

3.4.3 $\mathscr{L}_3$: Poisson likelihood for the total number of clusters

Finally, the third factor in our total likelihood, $\mathscr{L}_3$ (see Eq. 3.3), is $P(N)$, the probability of including a total of N clusters in our sample (the MMF3 cosmology sample). Since we assume that clusters are statistically independent from each other, $P(N)$ is simply a Poisson distribution with expected value $\bar{N}$, where $\bar{N}$ is given by Eq. (3.17), i.e.,

$$P(N) = \frac{\bar{N}^N e^{-\bar{N}}}{N!}. \tag{3.20}$$

3.4.4 Model parameters and priors

Cluster parameters

For convenience, we summarise below the set of cluster model parameters on which our likelihood, $\mathscr{L}$, depends:

$$\boldsymbol{p}_{\rm cl} = \{1 - b_{\rm SZ}, 1 - b_{\rm CMBlens}, \sigma_{\rm SZ}, \sigma_{\rm CMBlens}, r_{\rm CMBlens-SZ}, Y_\star, \alpha, \beta\}. \tag{3.21}$$

A list of these parameters can also be found in Table 3.1, along with their definitions.

We adopt priors on these cluster parameters in our likelihood analysis as follows. For the SZ parameters, we fix $\beta = 0.66$ (i.e., self-similar evolution), following the main analysis of Planck 2015 results XXIV (2016). Although the redshift evolution of the SZ scaling relation is not well constrained observationally, Planck 2015 results XXIV (2016) showed that the main impact of allowing β to vary within a broad Gaussian prior, $\beta = 0.66 \pm 0.50$, was to broaden constraints along the $\Omega_{\rm m}$–σ_8 degeneracy direction, rather than in the perpendicular direction that responds to the absolute cluster mass scale (see their Appendix A.3). We expect similar effects in our analysis, with only a small impact on constraints on the mass bias parameter $1 - b_{\rm SZ}$ of primary interest. We

Table 3.1 Summary of the cluster parameters of our model.

Parameter	Definition
$1 - b_{\rm SZ}$	SZ mass bias
$1 - b_{\rm CMBlens}$	CMB lensing mass bias
$\sigma_{\rm SZ}$	SZ intrinsic scatter
$\sigma_{\rm CMBlens}$	CMB lensing intrinsic scatter
$r_{\rm CMBlens\text{-}SZ}$	SZ–CMB lensing intrinsic correlation coefficient
$Y_\star$	Scaling relation normalisation
α	Scaling relation mass exponent
β	Scaling relation $E(z)$ exponent

also fix the intrinsic scatter in the SZ signal-to-noise to $\sigma_{\rm SZ} = 0.173$, the central value of the empirically-derived prior adopted in Planck 2015 results XXIV (2016), for reasons of computational efficiency.[6] For the remaining parameters in the SZ scaling relation, α and $\log Y_\star$, we adopt the Gaussian priors reported in Table 3.2, following Planck 2015 results XXIV (2016). Finally, for the SZ mass bias we adopt a flat positivity prior $1 - b_{\rm SZ} \geq 0$.

For the CMB lensing mass parameters, we impose Gaussian priors on $1 - b_{\rm CMBlens}$, $\sigma_{\rm CMBlens}$, and $r_{\rm CMBlens-SZ}$ with means and standard deviations given in Table 3.2. In Chapter 4, we study the intrinsic bias and scatter of our CMB lensing observable using N-body simulations (as was done, e.g., in Becker and Kravtsov (2011) for galaxy weak lensing). Our prior on $1 - b_{\rm CMBlens}$ is taken to be consistent with the values we find in Chapter 4 in addition to the bias expected from miscentering. Similarly, our prior on $\sigma_{\rm CMBlens}$ is taken to be consistent with the values we report in Chapter 4. Finally, our prior for the correlation between the intrinsic scatter on SZ and CMB lensing mass signal-to-noise ratios, $r_{\rm CMBlens-SZ}$, is taken to be consistent with the values of the SZ–lensing mass intrinsic correlation reported in Shirasaki et al. (2016).

Cosmological parameters

We also stress the dependence of our likelihood on cosmological parameters, so we can jointly constrain cosmology and cluster physics. The strongest dependences are on $\Omega_{\rm m}$,

[6] The constraints on this parameter from Planck 2015 results XXIV (2016) for the CCCP and WtG calibrations are prior-driven, the central values being a fraction of a σ away from the prior central value. In addition, we find from their corresponding MCMC chains that $\sigma_{\rm SZ}$ has small correlation with our main parameter of interest, $\sigma_8 \, (\Omega_{\rm m}/0.33)^{0.25}$, with a correlation coefficient of about 0.2 for both calibrations (see Section 3.6 for the motivation behind this parametrisation). We therefore conclude that letting this parameter vary and imposing on it the same prior as in Planck 2015 results XXIV (2016) would have little impact on our results.

Table 3.2 Means and standard deviations of the Gaussian priors adopted in our likelihood analysis of the real cluster data. Parameters not listed have broad flat priors.

Parameter	Mean	Standard deviation
$1 - b_{\rm CMBlens}$	0.93	0.05
$\sigma_{\rm CMBlens}$	0.20	0.05
$r_{\rm CMBlens\text{-}SZ}$	0.83	0.10
α	1.79	0.08
$\log Y_\star$	-0.19	0.02
$100\theta_{\rm MC}$	1.04093	0.00030

σ_8, and H_0 through the halo mass function, the volume element, and the mean values of the SZ and CMB lensing signal-to-noise ratios at given M_{500}, z, and sky location. There are only weak dependencies on the baryon density, $\Omega_{\rm b}h^2$, and spectral index of the primordial curvature perturbations, $n_{\rm s}$, through the shape of the matter power spectrum around cluster scales that enters the halo mass function. In our analysis, we fix these two parameters to the values $\Omega_{\rm b}h^2 = 0.0223$ and $n_{\rm s} = 0.9667$ determined by *Planck* (Planck 2015 results XIII, 2016). We also fix $\Omega_K = 0$ and the summed neutrino mass to the minimal value, $\sum m_\nu = 0.06\,{\rm eV}$.

We impose broad, flat priors on $\Omega_{\rm m}$, σ_8, and h. However, since h is poorly constrained by the cluster data, we further impose a Gaussian prior on the CMB acoustic scale parameter, $\theta_{\rm MC}$ (see Table 3.2). This parameter is an analytic approximation to the ratio of the sound horizon at CMB last scattering to the angular diameter distance – see Kosowsky et al. (2002), where $\theta_{\rm MC}$ is their $\mathscr{A}$ – which at fixed baryon density depends only on $\Omega_{\rm m}$ and H_0 in flat ΛCDM models. The acoustic scale parameter is very well measured by *Planck* from the acoustic peak locations, and is almost model independent. Our prior is from the analysis reported in Planck 2015 results XIII (2016). We stress that this prior is *geometric* and uses no information on the amplitude of the fluctuations from the CMB power spectrum.

3.5 Likelihood validation

We test the implementation of our likelihood by analysing a set of simulated data, $\{N, \hat{\underline{\boldsymbol{n}}}, \underline{D}\}$. We produce these data following our model assumptions at fixed values of cosmological and cluster model parameters, and then explore the corresponding posterior distribution with Markov-chain Monte Carlo (MCMC).

3.5.1 Simulated data

In order to generate our set of simulated data, we assume a flat ΛCDM cosmology with $\Omega_{\rm m} = 0.315$, $\sigma_8 = 0.811$, $h = 0.674$, and $\Sigma m_\nu = 0.06\,\mathrm{eV}$. We set the cluster model parameters close to the mean values of the Gaussian priors in Table 3.2 along with $\beta = 0.66$, $\sigma_{\rm SZ} = 0.173$ (as fixed by our priors), and $1 - b_{\rm SZ} = 0.62$. These fiducial values are listed in the final column of Table 3.3.

The first step in producing our simulated dataset is to obtain N, the total number of clusters that will be in our sample. We obtain it by drawing one sample from a Poisson distribution with mean value $\bar{N}$ given by Eq. (3.17), evaluated at our assumed values of cosmological and model parameters. We obtain a total of $N = 416$ clusters.

We then generate the set of sky locations of these N clusters, $\underline{\hat{\boldsymbol{n}}}$. To do this, in order to produce a catalogue that is statistically as close as possible to the MMF3 catalogue we use the SZ matched filter noise estimates across the sky produced by the *Planck* collaboration and that were used in the construction of the MMF3 catalogue. In these, the SZ matched filter noise is estimated for a set of filter angular scales in a set of patches that cover the whole sky. For each patch, interpolating over these tabulated values and using Eq. (3.17), where we restrict the integration over the sky to the extent of the patch, we compute the expected number of clusters falling within the patch. For patch i, this is $\bar{N}_i$, and the probability for a detected cluster to be in that patch is simply $P_i = \bar{N}_i/\bar{N}$. We then assign a number of clusters to each patch by obtaining N samples from a multinomial distribution with probabilities P_i.

Finally, we produce a mass data point, D_i, for each of the Nclusters. We generate each of them in the following way. First, we obtain values of M_{500} and z by drawing a sample from $P(M_{500}, z|\hat{\boldsymbol{n}})$ (see Eq. 3.15) using rejection sampling. We then produce values of $q_{\rm t}$, $p_{\rm t}$, and $z_{\rm t}$ by drawing one sample from Eq. (3.13), with the values of M_{500} and z obtained in the previous step and $\hat{\boldsymbol{n}}$ as conditioning information. In this step, for each cluster, $\bar{q}$ is computed with the SZ matched filter noise estimate of the sky patch in which the cluster is located. On the other hand, $\bar{p}$ is computed with the empirical average CMB lensing matched filter noise of all the clusters in the MMF3 cosmology sample falling within the cluster's sky patch. This is for the sake of simplicity, since in our CMB lensing mass pipeline we have produced such noise estimates only at the MMF3 cluster locations. To avoid inconsistencies in the analysis of the *simulated* data, we use these averaged CMB lensing noise estimates in the likelihood in this case (but not in the analysis of the real cluster sample). Finally, we generate values of $q_{\rm obs}$, $p_{\rm obs}$, and $z_{\rm obs}$ by drawing one sample from Eq. (3.14), with the values of $q_{\rm t}$, $p_{\rm t}$, and $z_{\rm t}$ obtained in the previous step as conditioning information. If $q_{\rm obs} > 6$, we set

$(q_{\rm obs}, p_{\rm obs}, z_{\rm obs})$ as the mass data point corresponding to the cluster at sky location $\hat{\boldsymbol{n}}$. If not, we repeat the process again until $q_{\rm obs} > 6$.

3.5.2 Parameter constraints

In order to validate the implementation of our likelihood, $\mathscr{L}$, we apply it to the simulated dataset, described in Section 3.5.1, and explore the corresponding posterior distribution with an MCMC. We allow the same set of cosmological and cluster model parameters to vary as in our analysis of the real data, $\boldsymbol{p} = \{\Omega_{\rm m}, \sigma_8, h, b_{\rm SZ}, b_{\rm CMBlens}, \sigma_{\rm CMBlens}, r_{\rm CMBlens-SZ}, \alpha, Y_\star\}$, and impose similar priors on them. For those parameters on which we impose Gaussian priors when analysing the real data (see Table 3.2), we retain the widths of the priors but, where necessary, recentre the means on the input values used to generate the simulations. As for the real data, we impose broad, flat priors on $\Omega_{\rm m}$, σ_8, h, and $1 - b_{\rm SZ}$, while $\sigma_{\rm SZ}$ and β are fixed to their input values.

We explore our posterior using the `emcee` package, which performs affine-invariant MCMC sampling.[7] The two-dimensional marginalised constraints that we obtain for $\Omega_{\rm m}$, σ_8, h, and $1 - b_{\rm SZ}$ are shown in Figure 3.2, and the one-dimensional marginalised constraints for all the parameters varied in the analysis are given in Table 3.3. It can be seen that these constraints are consistent with the input parameter values used to construct the simulated data. The two-dimensional marginalised constraints on the remaining parameters are not shown, but are similarly consistent with their true values. This validates our likelihood implementation and, as long as our Bayesian model remains a good description of the real data, our analysis. Of course, as our validation is limited to one simulation, we cannot test for systematic biases at a level below the statistical errors expected from our real data.

3.6 Results and discussion

We explore the posterior distribution for the real cluster data with an MCMC using the `emcee` package, as for our tests on simulated data. The two-dimensional marginalised constraints on all the parameters we allow to vary are shown in Figure 3.3, and the two-dimensional marginalised constraints on $\Omega_{\rm m}$, σ_8, and $1 - b_{\rm SZ}$ are shown in Figure 3.4, along with the corresponding constraints on the cosmological parameters from the *Planck* 2018 measurements of the CMB angular power spectra and CMB lensing (the TT,TE,EE+lowE+lensing likelihood; Planck 2018 results VI 2018). In

[7] http://dfm.io/emcee/current/

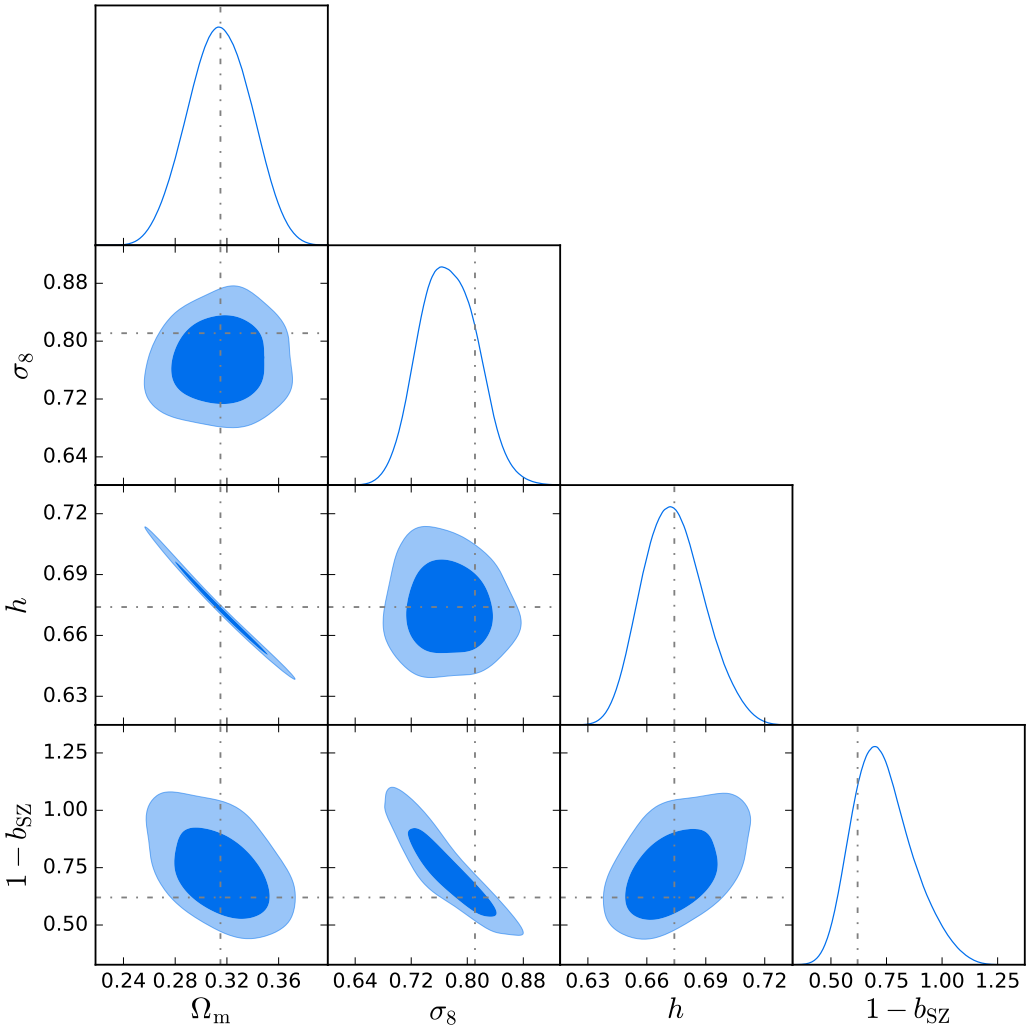


Fig. 3.2 Constraints (68 % and 95 % confidence regions) on $\Omega_{\rm m}$, σ_8, h, and $1 - b_{\rm SZ}$ for our simulated cluster sample. One-dimensional, marginalised posterior distributions are also shown. The input values of the parameters are shown as dash-dotted lines.

Table 3.3 Marginalised constraints (mean and 68 % confidence limits) on the parameters that we allow to vary in our analysis of the simulated data, along with their corresponding input values.

Parameter	Constraints (68 % uncert.)	Input value
Ω_m	0.31 ± 0.02	0.315
σ_8	0.77 ± 0.04	0.811
h	0.67 ± 0.02	0.674
$1 - b_\mathrm{SZ}$	0.72 ± 0.1	0.62
$1 - b_\mathrm{CMBlens}$	0.91 ± 0.05	0.92
σ_CMBlens	0.23 ± 0.05	0.22
$r_\mathrm{CMBlens\text{-}SZ}$	0.75 ± 0.09	0.77
α	1.84 ± 0.06	1.79
$Y_\star$	0.65 ± 0.03	0.646

addition, the one-dimensional marginalised constraints on $\sigma_8 \, (\Omega_\mathrm{m}/0.33)^{0.25}$ and on our cluster parameters are given in Table 3.4.

We note that the constraints on Ω_m, σ_8, and h are dependent on our choice of priors on α (Gaussian prior) and β (delta function prior). As argued in Section 3.4.4, allowing β to vary would widen the constraints in the Ω_m–σ_8 plane along the long degeneracy axis. Through the very strong degeneracy between Ω_m and h due to our prior on θ_MC, the constraints on h would also widen accordingly. Planck 2015 results XXIV (2016) find that leaving α free also widens the constraints along the long degeneracy axis in the Ω_m–σ_8 plane; we expect a similar effect in our analysis, with constraints on h also widening accordingly. We therefore decide to quote our cosmological constraints in terms of $\sigma_8 \, (\Omega_\mathrm{m}/0.33)^{0.25}$, a parameter that runs perpendicular to the long degeneracy axis in the Ω_m–σ_8 plane, and which therefore is more immune to the choice of priors on α and β. Purely for reference, we find $\Omega_\mathrm{m} = 0.33 \pm 0.02$, $\sigma_8 = 0.76 \pm 0.04$, and $h = 0.66 \pm 0.01$, stressing that they are dependent on the choice of priors on α and β.

As it can be seen in Table 3.4, our constraint on $\sigma_8 \, (\Omega_\mathrm{m}/0.33)^{0.25}$ from the *Planck* MMF3 cosmology sample of clusters with CMB lensing calibration of the cluster masses is consistent with that derived in the ΛCDM model from the *Planck* 2018 CMB angular power spectra (and lensing), $\sigma_8 \, (\Omega_\mathrm{m}/0.33)^{0.25} = 0.8020 \pm 0.0085$. Consistency can also be seen in Figure 3.4. We therefore find no evidence for tension between the abundance of galaxy clusters measured at lower redshifts and the predictions for the ΛCDM model with parameters (mostly) calibrated at CMB decoupling. An alternative view of this consistency is to compare our constraint on the SZ mass bias parameter, $1 - b_\mathrm{SZ} = 0.71 \pm 0.10$ (see Table 3.5), with that obtained in ΛCDM from joint analysis

Table 3.4 Marginalised constraints (mean and 68 % confidence limits) on $\sigma_8 \, (\Omega_\mathrm{m}/0.33)^{0.25}$ and on the cluster parameters that we allow to vary in our analysis of the *Planck* MMF3 cosmology sample of clusters.

Parameter	Constraints (68% uncert.)
$\sigma_8 \, (\Omega_\mathrm{m}/0.33)^{0.25}$	0.765 ± 0.035
$1 - b_\mathrm{SZ}$	0.71 ± 0.1
$1 - b_\mathrm{CMBlens}$	0.92 ± 0.05
σ_CMBlens	0.21 ± 0.04
$r_\mathrm{CMBlens\text{-}SZ}$	$0.80^{+0.06}_{-0.09}$
α	1.89 ± 0.07
$Y_\star$	0.65 ± 0.03

Table 3.5 Constraints (mean and 68 % confidence region) on $1 - b_\mathrm{SZ}$ from the datasets indicated. These include weak lensing of background galaxies for subsets of *Planck* clusters from Weighting the Giants (WtG; von der Linden et al. 2014), the Canadian Cluster Comparison Project (CCCP; Hoekstra et al. 2015), Sereno et al. (2017, the "cosmological subsample"), Penna-Lima et al. (2017), and Medezinski et al. (2018). CMB-derived constraints are for the CMB lensing calibration presented in Planck 2015 results XXIV (2016), the *Planck* 2015 SZ counts together with the *Planck* 2018 TT,TE,EE+lowE likelihood (Planck 2018 results VI, 2018), and this work. The constraint from the combined analysis of *Planck* clusters, the tSZ power spectrum, and the tSZ bispectrum of Hurier and Lacasa (2017) is also included.

Dataset	$1 - b_\mathrm{SZ}$
WtG	0.688 ± 0.072
CCCP	0.780 ± 0.092
Sereno et al. (2017)[a]	0.66 ± 0.10
Penna-Lima et al. (2017)	0.73 ± 0.10
Medezinski et al. (2018)	0.80 ± 0.14
Hurier et al. (2017)	0.71 ± 0.07
Planck 2015 CMB lensing[b]	$1.01^{+0.24}_{-0.16}$
Planck 2015 SZ + *Planck* 2018 CMB	0.62 ± 0.04
This work	0.71 ± 0.10

[a] In our notation, Sereno et al. (2017) actually constrain $\ln(1 - b_\mathrm{SZ}) = -0.40 \pm 0.14$.

[b] The CMB lensing measurement in Planck 2015 results XXIV (2016) is actually $1/(1 - b_\mathrm{SZ}) = 0.99 \pm 0.19$.

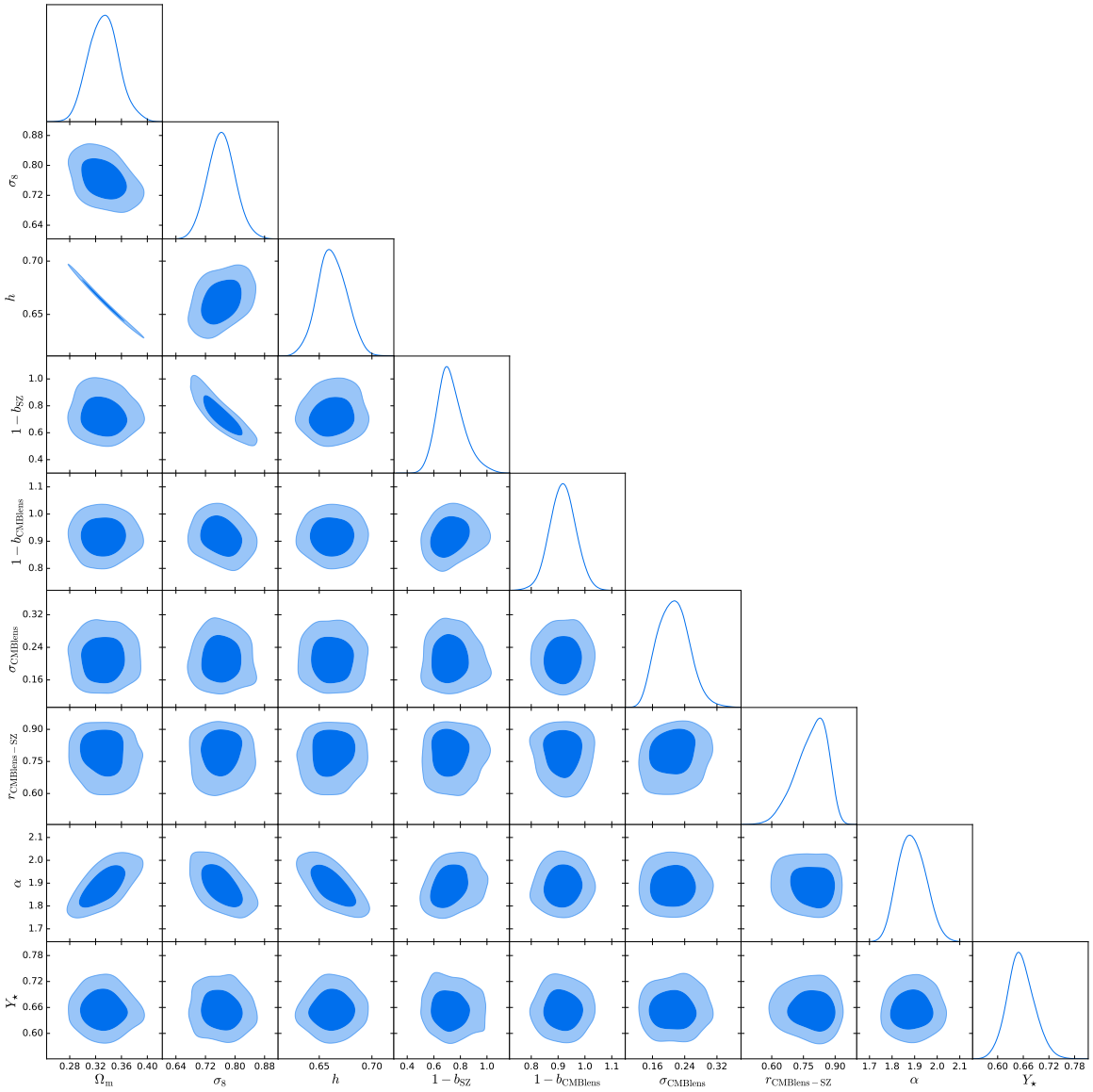


Fig. 3.3 Constraints (68 % and 95 % confidence regions) on the cosmological and cluster model parameters that we allow to vary in our analysis of the *Planck* MMF3 cosmology sample of clusters with CMB lensing mass calibration and a prior on θ_MC. One-dimensional, marginalised posterior distributions are also shown.

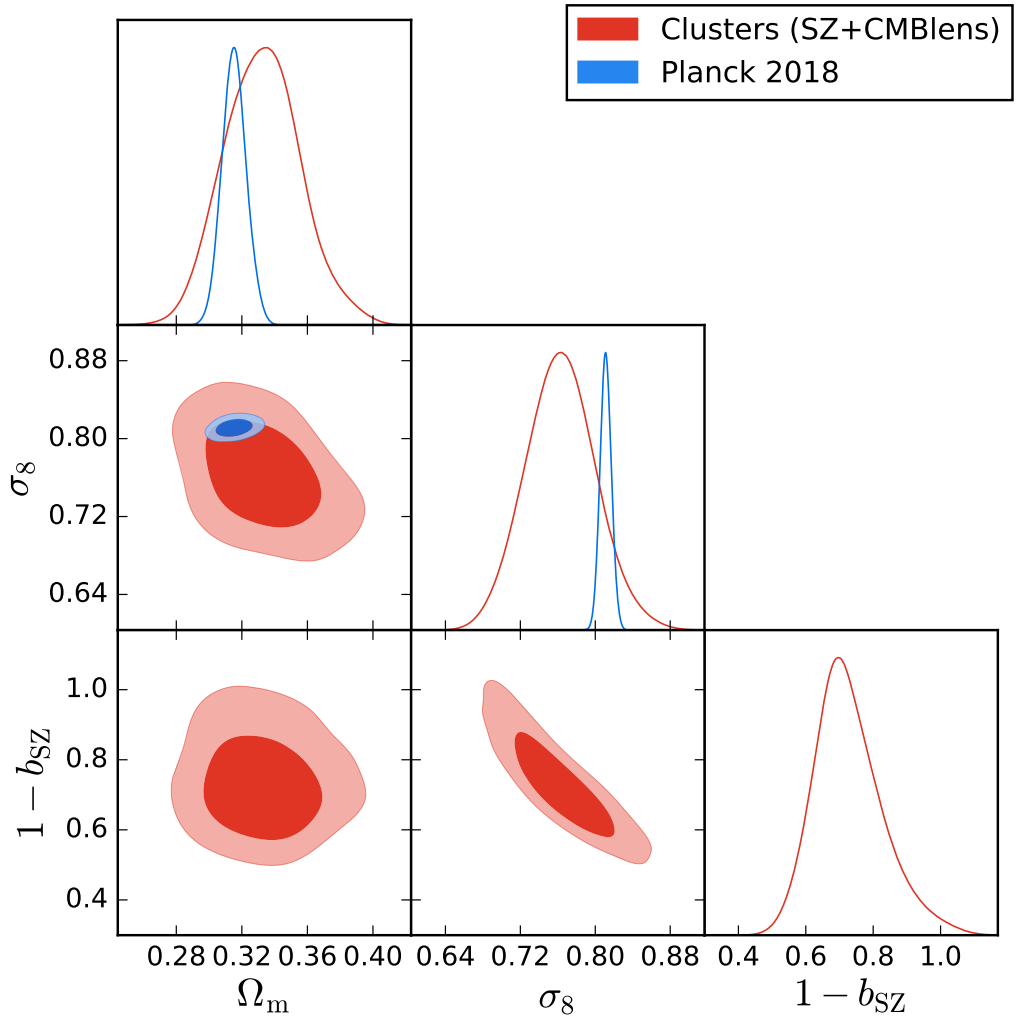


Fig. 3.4 Constraints on the cosmological parameters $\Omega_{\rm m}$ and σ_8, and on the SZ mass bias, $1-b_{\rm SZ}$, from our analysis of the *Planck* MMF3 cosmology sample of clusters (red). We stress that the joint constraints on $\Omega_{\rm m}$ and σ_8 along their long degeneracy axis are strongly dependent on the priors on α and β, whereas the perpendicular direction (the approximate combination $\sigma_8\Omega_{\rm m}^{0.25}$) is more immune to these priors. The one-dimensional constraints on $\Omega_{\rm m}$ and σ_8 are similarly strongly affected. Also shown (in blue) are the constraints from the *Planck* 2018 measurements of the CMB angular power spectra and CMB lensing, using the data combination TT,TE,EE+lowE+lensing (Planck 2018 results VI, 2018). Both constraints assume a flat ΛCDM cosmology.

of the CMB angular power spectra and cluster counts without further mass calibration. The latter was reported as $1 - b_{\rm SZ} = 0.62 \pm 0.04$ by Planck 2018 results VI (2018), showing good consistency.

Our CMB-lensing constraints on the SZ mass bias are also consistent with those from lensing of background galaxies by Weighing the Giants (WtG; von der Linden et al. 2014) and from the Canadian Cluster Comparison Project (CCCP; Hoekstra et al. 2015); see Table 3.5. These mass calibrations were used as priors in the analysis of Planck 2015 results XXIV (2016). They are also consistent with the weak-lensing constraints of subsamples of *Planck* clusters by Penna-Lima et al. (2017) and Sereno et al. (2017), with the constraint from the combined analysis of *Planck* clusters and the tSZ power spectrum and bispectrum of Hurier and Lacasa (2017), and with the HSC galaxy weak-lensing mass calibration of eight ACTPol clusters of Miyatake et al. (2019) and of five *Planck* clusters of Medezinski et al. (2018) (see also Table 3.5). We note also that our constraint on $1 - b_{\rm SZ}$ is of comparable precision to current constraints from galaxy lensing. While the mass constraints on individual low-redshift clusters are weaker from current CMB lensing data than from galaxy lensing, the former approach can be applied to the full cluster sample.

The general consistency between the constraints obtained in our analysis and those derived from the *Planck* CMB and from *Planck* SZ counts with the CCCP and the WtG mass calibrations is further illustrated in Figure 3.5, which shows the constraints in the $\Omega_{\rm m}$–σ_8 plane derived from these analyses, as well as those from the *Planck* SZ counts analysis with the CMB lensing calibration from Planck 2015 results XXIV (2016). In addition, our constraint on σ_8 is consistent with constrains on this parameter obtained in other recent cosmological analyses, as illustrated in Table 3.6.

Our constraint on the SZ mass bias is in mild disagreement with that obtained from CMB lensing in Planck 2015 results XXIV (2016). Reasons for this discrepancy were discussed in Section 3.3. Our constraint is, however, of similar precision: we find $1/(1 - b_{\rm SZ}) = 1.41 \pm 0.20$ from our samples of $b_{\rm SZ}$. We also note that our cluster CMB lensing measurements are consistent with those presented in Planck 2015 results XXIV (2016). Indeed, we can consider the CMB lensing mass measurements obtained with our matched filter (Eq. 2.23), as opposed to the signal-to-noise measurements used in our likelihood analysis, and use them to estimate $1/(1 - b_{\rm SZ})$ with the estimator employed in Planck 2015 results XXIV (2016), $\hat{1}/(1 - b_{\rm SZ}) = M_{\rm lens}/M_{Y_z}$. Although, as explained in Section 2.4.3, these mass measurements depend on the matched filter input angular scale, for each cluster we use the same input angular scale as in Planck 2015 results XXIV (2016), namely that given by the corresponding *Planck* SZ mass proxy, so the

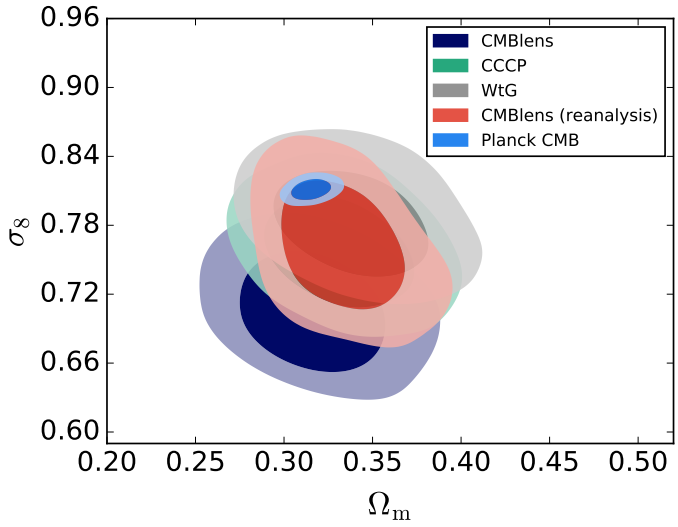


Fig. 3.5 Constraints in the Ω_m–σ_8 plane from our analysis (in red), along with those from the *Planck* 2015 SZ counts analysis with the CCCP, WtG, and CMB lensing mass calibrations (in green, grey, and dark blue, respectively), and those derived from the *Planck* 2018 CMB angular power spectra (in light blue).

Table 3.6 Constraints on σ_8 (mean and 68 % confidence region) from the indicated cosmological analyses.

Analysis	σ_8
Cluster counts	
This work	0.76 ± 0.04
SPT (SZ, Bocquet et al. 2019)	0.781 ± 0.037
Planck+CCCP (SZ, Planck 2015 results XXIV 2016)	0.76 ± 0.03
Planck+WtG (SZ, Planck 2015 results XXIV 2016)	0.78 ± 0.03
ACT (SZ, Hasselfield et al. 2013)	0.872 ± 0.065
XXL (X-ray, Pacaud et al. 2018)	0.721 ± 0.071
HIFLUGCS (X-ray, Schellenberger and Reiprich 2017)	0.79 ± 0.03
RASS+WtG (X-ray, Mantz et al. 2015)	0.83 ± 0.04
`GalWCat19` (optical, Abdullah et al. 2020)	$0.810^{+0.053}_{-0.056}$
SDSS (optical, Costanzi et al. 2019)	$0.91^{+0.11}_{-0.10}$
SDSS (optical, Kirby et al. 2019)	$0.85^{+0.06}_{-0.08}$
Cosmic shear and galaxy clustering	
DES-Y1 (Abbott et al., 2019)	$0.817^{+0.045}_{-0.056}$
KiDS+GAMA (van Uitert et al., 2018)	$0.776^{+0.064}_{-0.081}$
CMB	
Planck TT,TE,EE+lowE+lensing (Planck 2018 results VI, 2018)	0.8111 ± 0.0060

comparison should be valid. With this estimator, weighting each cluster by its inverse estimated variance, we find $\hat{1}/(1-b_{\rm SZ}) = 1.09 \pm 0.24$, which is consistent at $0.33\,\sigma$ with the measurement reported in Planck 2015 results XXIV (2016), $\hat{1}/(1-b_{\rm SZ}) = 0.99 \pm 0.19$.

There are some degeneracies between parameters that are worth commenting on. The strongest degeneracy is that between $\Omega_{\rm m}$ and h (see Figs. 3.3 and 3.4). This degeneracy is due to the prior we impose on $\theta_{\rm MC}$, which, at fixed baryon density, constrains $\Omega_{\rm m}h^3$ to be approximately constant. More interesting, regarding the cluster counts, is the degeneracy between $1-b_{\rm SZ}$ and σ_8. As expected, there is a negative correlation between these parameters, which arises mostly through the dependence of our likelihood on the halo mass function. The physical interpretation of this degeneracy is straightforward: a given set of cluster SZ measurements can be explained with some given values of $1-b_{\rm SZ}$ and σ_8, or with a smaller value of $1-b_{\rm SZ}$ (making the mass larger at a given SZ cluster signal) and a larger value of σ_8 (to enhance the number of more massive clusters to match the observed number). Finally, there is also an anti-correlation between $\Omega_{\rm m}$ and σ_8, arising primarily from the mass function: the increase in the number density of massive clusters with increasing σ_8 can be offset by the overall reduction in the number density of clusters with decreasing matter density.

Finally, we note that our CMB lensing signal-to-noise $p_{\rm obs}$ measurements are interesting per se in that they represent a significant detection of cluster CMB lensing with *Planck* data. Since the observational scatter is much larger than the intrinsic scatter and than the scatter associated with the spread of M_{500} and z across the sample, our $p_{\rm obs}$ measurements roughly follow a Gaussian distribution with some mean and standard deviation of unity. If no cluster lensing signal were present, they would follow a Gaussian distribution with zero mean and unit standard deviation. We find $\langle p_{\rm obs}\rangle = 0.234 \pm 0.052$, where angular brackets denote averaging over the cluster sample, thus detecting the CMB cluster lensing signal at $4.5\,\sigma$ significance. For comparison, Baxter et al. (2015) detect the CMB lensing cluster signal of 513 SZ-selected SPT clusters with SPT data at $3.1\,\sigma$ significance, Baxter et al. (2017) detect that of 3697 high-redshift DES clusters with SPT data at $8.1\,\sigma$ significance, and Raghunathan et al. (2019b) detect that of 4003 and 1741 DES clusters with SPTpol data at $8.7\,\sigma$ and $6.7\,\sigma$ significance, respectively. In addition, Madhavacheril et al. (2015) detect the CMB lensing signal of 12 000 optically-selected CMASS galaxies using ACT data at $3.2\,\sigma$ significance.

3.7 Conclusion

In this chapter we have presented constraints on the ΛCDM-model parameters $\Omega_{\rm m}$, σ_8, and H_0, and a CMB lensing calibration of the SZ mass bias, $1-b_{\rm SZ}$, obtained from 439 SZ-selected galaxy clusters from the *Planck* MMF3 cosmology sample. Our analysis revisits the *Planck* SZ counts analysis with CMB lensing mass calibration presented in Planck 2015 results XXIV (2016). The analysis there used cluster counts in the SZ signal-to-noise and redshift plane in order to constrain cosmological (and cluster model) parameters, imposing a prior on $1-b_{\rm SZ}$ derived from a CMB lensing mass calibration also presented in Planck 2015 results XXIV (2016). Such calibration found no evidence for a SZ mass bias, in contrast to galaxy weak-lensing calibrations on subsets of the cluster sample (e.g., von der Linden et al. 2014; Hoekstra et al. 2015) that favour $1-b_{\rm SZ}<1$ at more than $2\,\sigma$. Although the statistical significance of the difference in mass bias measured from CMB and galaxy lensing is relatively weak, given the large measurement errors for CMB lensing, adopting the CMB lensing calibration of Planck 2015 results XXIV (2016) exacerbates tension between cluster constraints in the σ_8–$\Omega_{\rm m}$ plane and those derived from the primary CMB anisotropies in ΛCDM.

We argued in Section 3.3 that there are several effects that may have led to bias in this previous mass calibration from CMB lensing. We have remeasured the cluster masses via CMB lensing, and included the signal-to-noises of these and the SZ measurements, along with the cluster redshifts, in a Bayesian analysis that naturally takes account of all significant effects that likely biased the analysis in Planck 2015 results XXIV (2016). This approach allows us to constrain jointly the cosmological parameters and the SZ mass bias (and other cluster model parameters) in an unbiased way, as demonstrated through simulated data in Section 3.5.

With our likelihood, and including priors on some of the cluster model parameters informed by results from numerical simulations (but, notably, with no prior on the SZ mass bias $1-b_{\rm SZ}$), we obtain constraints on $\Omega_{\rm m}$, σ_8, and H_0 that are consistent with the constraints in the ΛCDM model derived from the *Planck* measurements of the CMB power spectra (see Figs. 3.3 and 3.4, and Table 3.4). We measure a significant SZ mass bias, $1-b_{\rm SZ}=0.71\pm0.10$, consistent with measurements from galaxy weak lensing. We therefore find no evidence in our analysis of tensions between ΛCDM model parameters derived from the *Planck* cluster sample and the primary CMB anisotropies.

Our work is further evidence of the growing power of CMB lensing to calibrate the overall cluster mass scale in SZ counts analyses. It will be a particularly powerful approach for future high-resolution CMB experiments, such as CMB-S4, which should detect around 10^5 clusters through their SZ signatures (Abazajian et al., 2016). Indeed,

Constraints from *Planck* clusters with CMB lensing mass calibration

CMB lensing allows one to estimate cluster masses from CMB data alone (plus, in principle, redshift measurements of each cluster), which implies that the whole cluster sample can be used in the calibration analysis, rather than just a small subsample, which is currently the case for galaxy lensing calibrations. It is also not affected by the uncertainties in the photometric redshifts of the background galaxies, which are a limiting factor in galaxy lensing analyses, and, furthermore, allows for the determination of masses of high-redshift clusters, where galaxy lensing mass reconstructions perform badly due to the dearth of background galaxies. For future CMB experiments such as CMB-S4, the CMB lensing signal will be such that, in SZ counts analyses, the SZ–mass scaling relations will be able to be calibrated completely with CMB lensing masses alone to sub-percent accuracy (Louis and Alonso, 2017). In order to achieve this, however, work will be required in order to understand better the biases, intrinsic scatter, and correlations in the cluster observables, constraining them more accurately through numerical simulations and taking into account their dependence on mass and redshift.

Chapter 4

Quantifying the statistics of CMB lensing galaxy cluster mass measurements

4.1 Introduction

As the largest gravitationally-bound structures in the Universe, galaxy clusters are powerful cosmological probes (see Chapter 1). In particular, their abundance as a function of mass and redshift, as given by the halo mass function, depends on the assumed cosmological model and on the values of its parameters. Cosmological information can therefore be extracted from the observed galaxy cluster abundance in what is known as 'cluster counts' analyses. In a spatially-flat ΛCDM cosmology, this abundance is particularly sensitive to the mean matter density of the Universe, which can be parametrised by Ω_{m}, and to the amplitude of the matter perturbations, which can be characterised by σ_8, the root mean square of the linear density fluctuations smoothed on a scale of $8\,h^{-1}\,\mathrm{Mpc}$. Cluster masses are, however, not directly observable, so cluster counts studies must rely on one or several cluster observables known to scale with cluster mass and use them as 'mass proxies'. These observables can come from observations across different regions of the electromagnetic spectrum and include optical richness, X-ray flux, Sunyaev–Zel'dovich (SZ) flux, and lensing mass estimates, all of which are known to trace cluster masses. In recent years, studies using such observables have yielded competitive cosmological constraints (e.g., Mantz et al. 2010; Hasselfield et al. 2013; Planck 2015 results XXVII 2016; de Haan et al. 2016; Bocquet et al. 2019; Costanzi et al. 2019; Kirby et al. 2019; see also the analysis of Chapter

3, published in Zubeldia and Challinor 2019). In these analyses, scaling relations are needed in order to relate accurately the cluster observables to the cluster mass, since it is the dependence of the cluster abundance on the latter that can be theoretically predicted. These scaling relations need to be calibrated, and this is typically a difficult task which often brings in significant systematic uncertainty. Indeed, currently, the determination of cluster masses constitutes the largest source of uncertainty in cluster counts studies (see Pratt et al. 2019 for a recent review of cluster mass calibration).

Lensing observations can be very useful in this respect, since they provide almost bias-free estimates of cluster masses. In SZ counts analyses, in which an SZ-derived observable is used both to select the sample and as a *precise* (i.e., high signal-to-noise) cluster mass proxy for each cluster, the overall mass scale can be determined through *exact* (i.e., unbiased, or nearly unbiased) lensing mass estimates. These lensing-derived masses generally either have a significantly lower signal-to-noise ratio per cluster than the SZ observable, are much more costly to obtain, or a combination of both. These lensing mass estimates can be obtained from galaxy lensing observations, a well-established practice (e.g., as in two of the calibrations used in the baseline analysis in Planck 2015 results XXVII 2016, in de Haan et al. 2016, and in Bocquet et al. 2019), or from CMB observations, a more novel approach (as in the third calibration of the baseline analysis in Planck 2015 results XXVII 2016, and in Chapter 3).

First considered in Seljak and Zaldarriaga (2000), CMB lensing by clusters has attracted some interest in recent years. Several mass estimators have been developed (see, e.g., Yoo and Zaldarriaga 2008; Melin and Bartlett 2015; Raghunathan et al. 2017; Horowitz et al. 2019), and the CMB lensing signal of galaxy clusters and, in general, halos, has been detected to moderate-to-high statistical significance in various recent works (Baxter et al. 2015; Madhavacheril et al. 2015; Planck 2015 results XXVII 2016; Baxter et al. 2017; Raghunathan et al. 2019b,c; see also Chapters 2 and 3). CMB lensing cluster mass estimation has several intrinsic virtues, particularly in the context of SZ surveys. First, mass estimates can be obtained from the same data set as the SZ observable, and are relatively cheap to obtain, in contrast to most current galaxy-lensing-derived masses. This enables one to obtain CMB lensing mass estimates for the totality of the clusters in the SZ sample in a relatively straightforward way provided that redshift measurements of the clusters are available (Planck 2015 results XXVII 2016; see also Chapter 2), as opposed to most current galaxy lensing mass determinations, in which typically the masses of only a small subset of clusters in the SZ sample are obtained. Moreover, the signal-to-noise of CMB lensing mass estimates does not decrease strongly with redshift (see, e.g., Melin and Bartlett 2015), whereas

high-redshift galaxy lensing mass determinations can suffer from the lack of a sufficient number of background galaxies. In addition, the CMB lensing signal is not affected by the uncertainties in the photometric redshifts of the background galaxies, which are a limiting factor in galaxy lensing analyses. However, for an experiment like *Planck*, CMB lensing mass estimates are much noisier than state-of-the-art galaxy lensing mass estimates (e.g., von der Linden et al. 2014; Hoekstra et al. 2015), with typical signal-to-noise ratios of a fraction of unity for large clusters (see Chapter 3). Furthermore, being a more novel technique, cluster CMB lensing mass estimation methods have not been tested to the extent that galaxy lensing methods have (see, e.g., Raghunathan et al. 2017 for a quantification of several possible systematics that can affect CMB lensing cluster mass measurements).

A fundamental element of a cluster cosmological analysis that needs to be determined accurately in order for the analysis to deliver unbiased cosmological information is the relation between the mass observable (or observables) and the true cluster mass. This involves determining both the scaling relation that relates some mean value of the observable(s) with the cluster true mass, accounting for any possible biases, and the statistical scatter that exists around the mean value of the observable(s), which typically has a complex origin and is difficult to predict from first principles (Allen et al., 2011). More rigorously, what needs to be specified is the conditional probability density function followed by the mass observable(s), obs, conditioned on the value of the true cluster mass, M_{500}, $P(\mathrm{obs}|M_{500})$. We note that here and throughout we choose as a cluster's 'true mass' the mass within a sphere within which the mean density is 500 times the critical density at the cluster's redshift, and we denote it with M_{500}. We also note that this conditional probability density function can also be conditioned on other variables, such as redshift z; we omit these possible additional conditioning variables here for concision, but $P(\mathrm{obs}|M_{500})$ should be thought of as potentially having more conditioning variables implicit. If a functional form is assumed for $P(\mathrm{obs}|M_{500})$, its parameters can be self-calibrated with real data. However, doing this can have a negative impact on the statistical power of the cosmological analysis, especially if the calibration is carried out at the cosmological analysis level. An alternative approach is to use cosmological simulations in order to determine $P(\mathrm{obs}|M_{500})$, or at least to inform the likely range of parameter values of some assumed functional form. This also makes it possible to assess whether the chosen functional form is a good description of $P(\mathrm{obs}|M_{500})$. However, this approach is obviously limited by the accuracy of the simulations themselves, something that is difficult to quantify.

The distribution $P(\text{obs}|M_{500})$ has been widely studied in the literature from both simulations and observations for a number of cluster observables (e.g., Pratt et al. 2009; Becker and Kravtsov 2011; Angulo et al. 2012; Bahé et al. 2012; Rasia et al. 2012; Shirasaki et al. 2016; Geach and Peacock 2017). Nevertheless, it has never been studied for a CMB lensing mass observable. This is the aim of this chapter: to quantify $P(\text{obs}|M_{500})$, where obs is the CMB lensing mass observable that is used in Chapters 2 and 3, the CMB lensing signal-to-noise $p_{\rm obs}$. We follow the simulation approach, producing mock observations of $p_{\rm obs}$ for all the clusters with $M_{500} > 2 \times 10^{14} M_{\odot}$ in two snapshots of a large N-body cosmological simulation, BigMDPL (Klypin et al., 2016). We think of $p_{\rm obs}$ as consisting of three different contributions,

$$p_{\rm obs} = p_{\rm c} + \Delta p_{\rm u} + \Delta p_{\rm n} \,. \tag{4.1}$$

Here, $p_{\rm c}$ is the contribution to the lensing mass observable from the cluster itself, which includes the variation due to cluster triaxiality, and from large scale structure (LSS) correlated with the cluster. The other two terms, $\Delta p_{\rm u}$ and $\Delta p_{\rm n}$, are the contributions originating from LSS uncorrelated with the cluster and from lensing reconstruction noise, respectively. We follow two different approaches in order to analyse our mock observations. In the first approach, which we call our *deconvolution* approach, we treat both $\Delta p_{\rm u}$ and $\Delta p_{\rm n}$ as noise and then use our mock observations in order to characterise $p_{\rm c}$. In the second approach, which we call our *extrapolation* approach, we only treat $\Delta p_{\rm n}$ as noise, and then characterise $p \equiv p_{\rm c} + \Delta p_{\rm u}$ with our mock observations, extrapolating our results to the full line of sight from $z = 0$ back to CMB last scattering.

The main motivation for this work is to justify the choice of priors imposed on the parameters $1 - b_{\rm CMBlens}$ and $\sigma_{\rm CMBlens}$, which quantify, respectively, the bias and intrinsic scatter in the CMB lensing mass measurements, in the likelihood analysis of Chapter 3. In particular, this is done with our deconvolution approach, which is the one followed in the hierarchical model of Chapter 3. However, we also hope that our methods to study $P(\text{obs}|M_{500})$ and that our main insights may be useful for future cluster counts analyses that may use CMB lensing masses as a mass proxy, which otherwise will have to determine their appropriate conditional probability distribution $P(\text{obs}|M_{500})$. Throughout this chapter we consider an idealised experiment with specifications similar to those of *Planck*; however, we also briefly consider how some of the results change if a different experimental set-up is considered.

This chapter is organised as follows. First, in Section 4.2 we describe the cosmological simulations we use in order to obtain our mock observations. Then, in Section 4.3 we explain how we make such mock observations. Next, in Section 4.4 we present

our results, characterising $P(\mathrm{obs}|M_{500})$ for a *Planck*-like experiment. In this section, we first explain our model of the mock observations in Section 4.4.1, and then in Sections 4.4.2 and 4.4.3 we present our results for our deconvolution and extrapolation approaches, respectively. In Section 4.5 we consider how the extrapolation approach results change if different experiment specifications are considered, and we finally conclude in Section 4.6.

4.2 Simulation and convergence maps

In our study we use two snapshots of the BigMDPL simulation, a large, state-of-the-art simulation part of the publicly-available MultiDark simulation suite (Klypin et al., 2016). It consists of 3840^3 particles of mass $2.4 \times 10^{10} h^{-1} M_\odot$ in a simulation box of $2.5 h^{-1}\,\mathrm{Gpc}$ with periodic boundary conditions and evolved in a flat ΛCDM cosmology with $H_0 = 67.8\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$, $\Omega_\mathrm{m} = 0.307$, baryon density parameter $\Omega_\mathrm{b} = 0.048$, $\sigma_8 = 0.829$, and scalar spectral index $n_s = 0.96$. We consider the snapshots at $z = 0.23$ and $z = 0.52$.

We use the halo positions as determined with the BDM ('Bound Density Maximum') technique (Klypin and Holtzman, 1997; Riebe et al., 2013) that are publicly available in the CosmoSim database[1]. We then measure the spherical overdensity mass of each halo, M_{500}, using as density contrast $\Delta = 500$ defined with respect to the critical density. We note that we do not remove unbound particles in this process. We refer to M_{500} as the halo (or cluster) 'true mass'. We only consider clusters with $M_{500} > 2 \times 10^{14} M_\odot$ and that are not subhalos of another halo, which yields a cluster catalogue with a total of 60 391 clusters in the $z = 0.23$ snapshot and of 29 561 clusters in the $z = 0.52$ snapshot. This cluster range corresponds to the most massive clusters of the Universe and spans most of the clusters in the *Planck* MMF3 cosmology sample, the sample used in the cosmological analysis in Planck 2015 results XXVII (2016) and in Chapter 3; see Planck 2015 results XXVII (2016) for how this sample is constructed.

In order to obtain our CMB lensing measurements, the key quantity that needs to be produced is the lensing convergence around the location of each cluster in our catalogue and along a given direction of observation. We generate convergence maps using the Born approximation; under this approximation, for a spatially-flat cosmology the lensing convergence can be written as

$$\kappa(\hat{\boldsymbol{n}}) = \frac{3}{2}\left(\frac{H_0}{c}\right)^2 \Omega_\mathrm{m} \int_0^{\chi_\star} \frac{\chi\,(\chi_\star - \chi)}{\chi_\star} \frac{\delta}{a} d\chi\,, \tag{4.2}$$

[1] `www.cosmosim.org`

where δ is the matter density contrast, a is the scale factor normalised to unity today, $\chi_\star$ is the comoving distance to last scattering, and H_0 is the Hubble constant.

We produce square convergence maps centred at the location of each cluster in our catalogue for both snapshots. Each map has an angular size of 128 arcmin along both directions perpendicular to the line of sight, and we always take the z-direction, as defined within the simulation, as our line-of-sight direction. Obviously, with our simulation snapshots we cannot compute the total lensing convergence from $\chi = 0$ all the way back to last scattering, $\chi = \chi_\star$, as in Eq. (4.2). We instead restrict our integration limits to a relatively small interval of comoving distance. Specifically, we take each cluster to be at the comoving distance given by the redshift of the snapshot to which it belongs, χ_c, and we then consider seven sets of integration limits in Eq. (4.2), all of them centred at χ_c but with an increasing comoving length χ_l, $[\chi_c - \chi_l/2, \chi_c + \chi_l/2]$. We take χ_l, which hereafter we will refer to as 'integration length', to be logarithmic spaced between 5 Mpc and 400 Mpc, thus taking values of 5, 10.4, 21.5, 44.7, 92.8, 192.7, and 400 Mpc (comoving). We do this in order to investigate how the statistics of our CMB lensing mass observable depend on the relative amount of large scale structure (LSS) correlated and uncorrelated with the cluster that is present in the convergence maps. We note that the way in which we compute our convergence maps from the simulation snapshots neglects the temporal evolution of δ along the line of sight, since in a snapshot all the particle positions are given at the same time. The error we introduce by doing this should be negligible for the relatively small integration lengths (on cosmic scales) that we are considering. We do, however, take into account the change of the geometrical lensing kernel in Eq. (4.2) along the line of sight.

To summarise, for each cluster in our catalogue at any of the two considered redshifts we compute seven convergence maps, each of which corresponds to a different integration length. We refer to these maps as 'cluster convergence maps' at redshift z with integration length χ_l.

Furthermore, for each of the two snapshots we compute 10^5 square convergence maps centred at random locations in the simulation box, with the same angular size as the cluster convergence maps (128 arcmin across each direction), and with the two largest integration lengths used in the computation of the cluster convergence maps. We refer to these maps as 'random convergence maps' at redshift z with integration length χ_l, and describe their interpretation and use in Section 4.4.

4.3 Mock observations

In this section we describe our CMB lensing cluster mass observable and how we obtain mock measurements from our cluster convergence maps in order to produce a mock data set, which we then analyse in Section 4.4.

We think of each cluster convergence map with a given integration length as the 'true' convergence map around the location of that cluster due to the mass distribution contained within the considered integration length, and we denote it with $\kappa_{\rm c}$. As mentioned in Chapter 1, a fixed lensing convergence remaps the CMB anisotropies in a way that makes it possible for it to be estimated (or 'reconstructed') from CMB observations alone. Throughout this work, except in Section 4.5, we consider an idealised *Planck*-like CMB experiment with Gaussian beam with full-width at half maximum (FWHM) of 5 arcmin and with Gaussian instrumental noise with temperature noise levels of 45 μK arcmin. We do not study the impact of uncleaned or residual foregrounds on our results. As already noted above, in Section 4.5 we investigate how some of our results change with different experimental specifications.

We consider the simplest of the CMB lensing reconstruction techniques, the quadratic estimators (Hu, 2001; Hu and Okamoto, 2002), thus called because they are quadratic in the CMB fields. Specifically, we only use the TT quadratic estimator, which produces a reconstructed convergence map out of two copies of a CMB temperature map. We make this choice because this estimator is close to the optimal lensing estimator for a *Planck*-like experiment. Using the flat-sky approximation, which is accurate given the small angular size of our convergence maps, the reconstructed convergence map that corresponds to one of our cluster convergence maps, $\hat{\kappa}$, can be written in Fourier space as

$$\hat{\kappa}(\mathbf{L}) = \kappa_{\rm c}(\mathbf{L}) + n(\mathbf{L})\,. \tag{4.3}$$

Here, $\kappa_{\rm c}(\mathbf{L})$ is the Fourier transform of the 'true' convergence (the cluster convergence map), and $n(\mathbf{L})$ is the Fourier transform of the reconstruction noise, which in our simulated experiment has contributions from the primary CMB fluctuations and from instrumental noise. This noise is approximately Gaussian (Lewis and Challinor, 2006). Following Hu et al. (2007), in our quadratic estimator implementation we impose a Fourier-space top-hat low-pass filter on the gradient leg of the estimator, zeroing the gradient map for modes with $L > 2000$. This ensures that the quadratic estimator remains unbiased around regions of large convergence, e.g., in the central regions of galaxy clusters. In addition, as is customary, we use the lensed CMB power spectrum in the weighting of the gradient leg of the estimator and in the normalisation, rather

than the unlensed CMB power spectrum, as this gives (approximately) the correct response of the estimator $\hat{\kappa}(\boldsymbol{L})$ to lenses at wavevector $\boldsymbol{L}$ averaged over all other lensing modes (Lewis et al., 2011; Hanson et al., 2011). This is the same form of the quadratic estimator that we used in our analysis of real *Planck* data in Chapter 2. It provides, to a good approximation, unbiased reconstructed convergence maps, i.e., $\langle\hat{\kappa}(\mathbf{L})\rangle = \kappa_{\rm c}(\mathbf{L})$, where angular brackets denote ensemble averaging over reconstruction noise, that is, over CMB and instrumental noise fluctuations.

In our study we do not produce reconstructed maps from our cluster convergence maps, $\kappa_{\rm c}$, but instead we directly use the cluster convergence maps themselves. The reasons for doing this are the following. First, as just noted, to a good approximation our lensing reconstruction method produces unbiased convergence estimates for the experimental specifications considered. In addition, the scatter of the final lensing mass observable arising from reconstruction noise is already well understood. Indeed, since to a good approximation the reconstruction noise is Gaussian, it is fully described by its variance, which can be written as

$$\left\langle \hat{\kappa}\left(\boldsymbol{L}\right) \hat{\kappa}^{*}\left(\boldsymbol{L}'\right)\right\rangle = \delta^{(2)}\left(\boldsymbol{L}-\boldsymbol{L}'\right) N_{\kappa}(\boldsymbol{L})\,, \tag{4.4}$$

where $N_{\kappa}(\boldsymbol{L})$ is well approximated in our case by $N^{(0)}(L)$, the normalisation of the quadratic estimator (see e.g., Hu 2001 for an analytic expression). As described below, our mass observable is linear in the convergence; thus, its scatter due to reconstruction noise is also Gaussian to a good approximation.

The next step of our measurement pipeline is to match-filter our cluster convergence maps with a cluster convergence model in order to obtain a mass observable measurement for each cluster. We use exactly the same matched-filter implementation that was used in Chapter 2, and which follows Melin and Bartlett (2015). As a cluster model we adopt a truncated NFW profile (Navarro et al., 1997)

$$\rho(r) = \begin{cases} \frac{\rho_0}{(r/r_s)(1+r/r_s)^2} & \text{if } r \leq R_{\rm trunc}\,, \\ 0 & \text{if } r > R_{\rm trunc}\,, \end{cases} \tag{4.5}$$

where r_s is a characteristic scale radius, $R_{\rm trunc}$ is the truncation radius, and ρ_0 is a characteristic density, given by

$$\rho_0 = \rho_c(z)\frac{500}{3}\frac{c_{500}^3}{\ln(1+c_{500}) - c_{500}/(1+c_{500})}\,, \tag{4.6}$$

where $\rho_c(z)$ is the critical density at the cluster's redshift, and c_{500} is the concentration parameter, which is defined as $c_{500} = R_{500}/r_s$. As in Chapter 2, we choose $R_{\rm trunc} = 5R_{500}$ and we fix $c_{500} = 3$. Thus, in our model, a cluster is completely specified by two parameters, e.g., M_{500} and redshift, z. We use this parameterisation throughout this chapter, and refer to M_{500} and z as the 'true cluster parameters'.

As detailed in Chapter 2, we can then write the convergence due to our cluster model at an angular separation from the cluster centre $\boldsymbol{\theta}$ as

$$\kappa_{\rm m}(\boldsymbol{\theta}) = \kappa_0 \kappa_{\rm t}(\boldsymbol{\theta}; \theta_s)\,. \tag{4.7}$$

Here, κ_0 is a normalisation such that $\kappa_0\theta_s^2$ is proportional to M_{500} and $\kappa_{\rm t}(\boldsymbol{\theta};\theta_s)$ is a circularly-symmetric template function that depends only on θ/θ_s, where θ_s is the angular size of the scale radius r_s, $\theta_s = r_s/d_A(z)$, with $d_A(z)$ the angular diameter distance to the cluster. The template is normalised to unity at $\theta = \theta_s$. Given an estimate of the convergence around a cluster, $\hat{\kappa}(\boldsymbol{L})$, and a fiducial value of the cluster angular size, $\theta_s^{\rm fid}$, an estimator for κ_0 can be written as

$$\hat{\kappa}_0 = \left[\int \frac{d^2\boldsymbol{L}}{2\pi} \frac{|\kappa_{\rm t}(\boldsymbol{L})|^2}{N_\kappa(\boldsymbol{L})}\right]^{-1} \int \frac{d^2\boldsymbol{L}}{2\pi} \frac{\hat{\kappa}(\boldsymbol{L})\kappa_{\rm t}^*(\boldsymbol{L})}{N_\kappa(\boldsymbol{L})}\,, \tag{4.8}$$

where $N_\kappa(\boldsymbol{L})$ is the variance of the reconstruction noise of $\hat{\kappa}(\boldsymbol{L})$, and where the dependence of $\kappa_{\rm t}$ on $\theta_s^{\rm fid}$ is left implicit. This inverse-variance weighting ensures that the estimator is minimum-variance, which can be written as

$$\sigma^2_{\kappa_0} \equiv \left\langle (\hat{\kappa}_0 - \kappa_0)^2 \right\rangle = \frac{1}{2\pi}\left[\int \frac{d^2\boldsymbol{L}}{2\pi} \frac{|\kappa_{\rm t}(\boldsymbol{L})|^2}{N_\kappa(\boldsymbol{L})}\right]^{-1}\,. \tag{4.9}$$

We use each cluster's true mass, M_{500}, as the model fiducial mass, from which we derive its fiducial angular scale $\theta_s^{\rm fid}$:

$$M^{\rm fid} = 500\frac{4\pi}{3}\left[d_A(z)c_{500}\theta_s^{\rm fid}\right]^3 \rho_c(z)\,. \tag{4.10}$$

As discussed in Chapter 2, from this estimator $\hat{\kappa}_0$ a mass estimator $\hat{M}_{500}$ can be obtained in a straightforward way, but the signal-to-noise on $\hat{\kappa}_0$ (or, equivalently, on $\hat{M}_{500}$) turns out to have better properties as an observable. This is defined as

$$p_{\rm obs} \equiv \hat{\kappa}_0/\sigma_{\kappa_0}\,. \tag{4.11}$$

As can be seen from Eq. (4.8), for a given cluster $p_{\rm obs}$ is an unbiased estimator of the cluster CMB lensing signal-to-noise only if the convergence model, $\kappa_{\rm m}$, is equal to the true cluster convergence, $\kappa_{\rm c}$. As shown in Chapter 2, $p_{\rm obs}$ is more immune to mismatch between $\kappa_{\rm m}$ and $\kappa_{\rm c}$. In particular, it is much less dependent on the choice of $\theta_s^{\rm fid}$. We therefore use $p_{\rm obs}$ as our cluster observable; $p_{\rm obs}$ is also the observable used in Chapter 2.

As mentioned above, the statistics of the lensing reconstruction noise are well understood, and due to this we do not produce reconstructed convergence maps of each cluster convergence map, but instead we use each cluster's true convergence, $\kappa_{\rm c}$ in place of $\hat{\kappa}$ in the matched-filtering process. That is, what we compute is the mean CMB lensing signal-to-noise, $p \equiv \langle p_{\rm obs} \rangle$, where angular brackets denote averaging over reconstruction noise, $n(\mathbf{L})$, for each cluster in our catalogue and for all our seven integration lengths:

$$p = \sqrt{2\pi} \left[\int \frac{d^2\boldsymbol{L}}{2\pi} \frac{|\kappa_{\rm t}(\boldsymbol{L})|^2}{N_\kappa(\boldsymbol{L})} \right]^{-1/2} \int \frac{d^2\boldsymbol{L}}{2\pi} \frac{\kappa_{\rm c}(\boldsymbol{L}) \kappa_{\rm t}^*(\boldsymbol{L})}{N_\kappa(\boldsymbol{L})} \,. \tag{4.12}$$

In part of our analysis – specifically, in Section 4.4.2 – we also treat as noise the contribution to the cluster convergence maps $\kappa_{\rm c}$ coming from large scale structure (LSS) uncorrelated with the cluster. In this case, the variance of the reconstructed convergence, after averaging over reconstruction noise and this uncorrelated LSS, is therefore $N^{(0)}(L) + C_L^{\kappa\kappa}$, where $C_L^{\kappa\kappa}$ is the convergence power spectrum. We also make measurements of p for all our cluster convergence fields using this variance in the matched filter.

In summary, we obtain two sets of mock measurements of our cluster CMB lensing mass observable, p, for all our clusters, which are at two different redshifts, and for the seven integration lengths χ_l considered. Each set has a different understanding of what is thought of as noise in the matched-filtering process. In the first set, noise is understood as being solely due to reconstruction noise; this data set, which we refer to as cluster data set B, is analysed in Section 4.4.3. In the second case, the variation due to uncorrelated LSS is also included in the noise budget; this data set, which we refer to as cluster data set A, is studied in Section 4.4.2.

We also apply our measurement pipeline to our random convergence maps. The idea is to obtain the response of our measurement method to an observation that is not centred on a cluster, but where only the convergence due to random LSS is present. The use of these observations will become clear in Section 4.4. For each random convergence map, we obtain a set of measurements following our measurement

pipeline, using as matched-filter fiducial mass $M^{\rm fid}$ the centres of 90 equally-spaced subdivisions of the interval $(2\text{–}20)\times10^{14}M_\odot$, and as redshift the corresponding snapshot redshift. Our 10^5 random convergence fields at each snapshot redshift yield 10^5 mock measurements for each mass bin, redshift, and integration length. They constitute two sets of measurements of p as a function of filter fiducial mass and integration length, which we refer to as random data set A (in which the LSS power spectrum is included in the matched filter inverse-variance weighting) and B (in which only the reconstruction noise power spectrum is included in the matched filter inverse-variance weighting).

4.4 Statistics of the CMB lensing cluster mass observable

In this section we analyse our mock observations, which we obtain as detailed in Section 4.3, in order to characterise the statistics of our CMB lensing cluster mass observable. First, in Section 4.4.1, we describe how we understand our observations in terms of random variables. We then study our mock observations following the deconvolution and extrapolation approaches in Sections 4.4.2 and 4.4.3, respectively.

4.4.1 Model of observations

Let us consider our *noisy* CMB lensing cluster mass observable, $p_{\rm obs}$, with either of our two choices of matched-filter inverse-variance weighting, and for a given integration length χ_l. Let us also consider a point in true cluster parameter space, M_{500}–z. We can think of $p_{\rm obs}$ as a random variable, with variability arising from different sources. First, there is reconstruction noise, which arises from the CMB and instrumental noise fluctuations. In addition, there is a contribution coming from variation in the true lensing convergence. Indeed, clusters with a given true mass M_{500} and at a given redshift z yield, in general, different lensing convergences: they have different shapes, being in general triaxial, and different large scale structure correlated and uncorrelated with them, which, in projection along the line of sight, also contributes to the convergence. Since $p_{\rm obs}$ is linear in the lensing convergence, this variability in the lensing convergence translates directly into variability in $p_{\rm obs}$.

We can write $p_{\rm obs}$ at a given point in true cluster parameter space, M_{500}–z, as a sum of three random variables (as in Eq. 4.1, repeated here for convenience),

$$p_{\rm obs} = p_{\rm c} + \Delta p_{\rm u} + \Delta p_{\rm n} \,. \tag{4.13}$$

Here, $\Delta p_{\rm n}$ denotes the contribution to $p_{\rm obs}$ coming from the lensing reconstruction noise. It is the random variable associated with the response of our matched filter at the given true cluster parameters to the reconstruction noise n (see Eq. 4.3). Next, $\Delta p_{\rm u}$ denotes the contribution coming from LSS uncorrelated with the cluster (hereafter and for concision, 'uncorrelated LSS'). It is the random variable associated with the response of our matched filter at the given true cluster parameters to our random convergence maps with the considered integration length χ_l. Finally, $p_{\rm c}$ denotes the contribution to $p_{\rm obs}$ arising from the cluster itself (that is, from cluster morphology and orientation) *and* from LSS along the line of sight correlated with the cluster (hereafter, 'correlated LSS'). It is defined as the variable that arises from subtracting $\Delta p_{\rm u}$ from p, the variable that results after averaging $p_{\rm obs}$ over $\Delta p_{\rm n}$. In practice, in our deconvolution approach we determine the distribution of p_c by deconvolving the distributions of p and $\Delta p_{\rm u}$ estimated from our mock data sets.

If some real $p_{\rm obs}$ measurements are to be used in a cosmological analysis (e.g., in order to determine the mass scale of the cluster sample), the conditional probability density function (pdf) followed by $p_{\rm obs}$ at the true cluster parameters, $P(p_{\rm obs}|M_{500}, z)$, needs to be determined. In this chapter we propose two different approaches in order to characterise this conditional pdf. In the first approach, which we refer to as the *deconvolution* approach and which is developed in Section 4.4.2, we treat the scatter arising from reconstruction noise ($\Delta p_{\rm n}$) and from uncorrelated LSS ($\Delta p_{\rm u}$) as observational noise, and we then characterise the signal arising only from the cluster itself and from correlated LSS ($p_{\rm c}$) with our mock observations. This was the approach underlying the hierarchical model of Chapter 3; here we aim to justify the choices of priors on the CMB lensing bias and intrinsic scatter parameters of that work, $1-b_{\rm CMBlens}$ and $\sigma_{\rm CMBlens}$, respectively. In the second approach, which we refer to as the *extrapolation* approach and which is studied in Section 4.4.3, we think of observational noise as being solely due to reconstruction noise ($\Delta p_{\rm n}$), and then characterise the signal due to the cluster itself and to both correlated and uncorrelated LSS ($p_{\rm c} + \Delta p_{\rm u}$), extrapolating the results to the full line of sight from $z = 0$ to last scattering.

4.4.2 Deconvolution approach

Method

In this approach, our matched filter has both the reconstruction noise and the lensing convergence power spectra in the inverse-variance weighting (see Eq. 4.8). Let us first think of a hypothetical set of simulated observations in which the variation due to lensing by uncorrelated LSS is present from $z = 0$ back to last scattering, and not just within a box of a given length along the line of sight, and where reconstruction noise is also present. In order to study $P(p_{\rm obs}|M_{500}, z; \chi_\star)$, we divide it into two layers[2].

The first layer is $P(p_{\rm obs}|p_{\rm c}, M_{500}, z; \chi_\star)$, which is approximately a Gaussian distribution centred on $p_{\rm c}$ with unit standard deviation. Indeed, first, both $\Delta p_{\rm u}$ and $\Delta p_{\rm n}$ have zero expected values. We empirically check that $\Delta p_{\rm u}$ has an expected value consistent with zero as a function of M_{500} with our random data sets, which can be thought of as realisations of $\Delta p_{\rm u}$ at a given integration length (see below). In addition, reconstruction noise is approximately Gaussian, and therefore so is $\Delta p_{\rm n}$, which is linear in it. As we show below, for our experimental specifications, the standard deviation of $\Delta p_{\rm n}$, $\sigma_{\Delta p_{\rm n}}$, is significantly larger than the standard deviation of $\Delta p_{\rm u}$, $\sigma_{\Delta p_{\rm u}}$, across the mass range considered, and so the variance of $\Delta p_{\rm u} + \Delta p_{\rm n}$ is dominated by the reconstruction noise. For modest levels of non-Gaussianity of $\Delta p_{\rm u}$ – expected given the long integration length $\chi_\star$, the relatively high redshift of the lenses involved, and the effective low-pass filtering due to the inverse-variance weighting in the matched filter – we therefore expect the distribution of $\Delta p_{\rm u} + \Delta p_{\rm n}$ to be close to Gaussian. We can calculate the variance of $\Delta p_{\rm u}$ by taking the second moment of Eq. (4.12) to find

$$\sigma^2_{\Delta p_{\rm u}} = \left[\int \frac{d^2\boldsymbol{L}}{2\pi} \frac{|\kappa_{\rm t}(\boldsymbol{L})|^2}{N_\kappa(\boldsymbol{L})}\right]^{-1} \int \frac{d^2\boldsymbol{L}}{2\pi} \frac{|\kappa_{\rm t}(\boldsymbol{L})|^2 C_L^{\kappa\kappa}}{N_\kappa(\boldsymbol{L})}\,, \tag{4.14}$$

where $C_L^{\kappa\kappa}$ is the lensing convergence power spectrum, and where, recall, $N_\kappa(\boldsymbol{L}) = N^{(0)}(L) + C_L^{\kappa\kappa}$. For $z = 0.23$ we find $\sigma_{\Delta p_{\rm u}}/\sigma_{\Delta p_{\rm n}} = 0.19$ for $M_{500} = 2 \times 10^{14} M_\odot$, and $\sigma_{\Delta p_{\rm u}}/\sigma_{\Delta p_{\rm n}} = 0.22$ for $M_{500} = 10^{15} M_\odot$; for $z = 0.52$ we find $\sigma_{\Delta p_{\rm u}}/\sigma_{\Delta p_{\rm n}} = 0.15$ for $M_{500} = 2 \times 10^{14} M_\odot$, and $\sigma_{\Delta p_{\rm u}}/\sigma_{\Delta p_{\rm n}} = 0.18$ for $M_{500} = 10^{15} M_\odot$. Finally, our matched-filter inverse-variance weighting guarantees that $\Delta p_{\rm u} + \Delta p_{\rm n}$ has unit standard deviation and hence so does the distribution $P(p_{\rm obs}|p_{\rm c}, M_{500}, z; \chi_\star)$. In the context of this *deconvolution* approach, we refer to $P(p_{\rm obs}|p_{\rm c}, M_{500}, z; \chi_\star)$ as *observational* scatter.

[2]Here, $\chi_\star$ denotes that the integral along the line of sight is performed from $\chi = 0$ to $\chi = \chi_\star$ (i.e., to last scattering); a semicolon is used to stress the fact that, unlike the other two conditioning variables, $\chi_\star$ is not a cluster-related variable.

The other layer in which we decompose the scatter of our observable is $P(p_{\rm c}|M_{500}, z; \chi_\star)$, Within the context of this approach, we refer to this variability as *intrinsic* scatter. We study this conditional distribution with our mock observations. First, we assume that $P(p_{\rm c}|M_{500}, z; \chi_l = 400\,{\rm Mpc})$ is, to a good approximation, equal to $P(p_{\rm c}|M_{500}, z; \chi_\star)$; we remind that $\chi_l = 400\,{\rm Mpc}$ is the largest integration length that we have considered. Indeed, beyond 200 Mpc from the cluster centre, the contribution from correlated LSS to the convergence should be negligible (see Section 4.4.3 for a quantitative discussion). Thus, $P(p_{\rm c}|M_{500}, z; \chi_l = 400\,{\rm Mpc})$ is what needs to be determined. In the following, we will denote this pdf simply with $P(p_{\rm c}|M_{500}, z)$.

Unfortunately, our mock observations do not provide us with samples from $P(p_{\rm c}|M_{500}, z)$. However, our cluster data set A can be thought of as consisting of samples from $P(p, M_{500}|z; \chi_l)$, where $p = p_{\rm c} + \Delta p_{\rm u}$, and where z can be either of our two redshifts ($z = 0.23$ and $z = 0.52$), and χ_l either of our seven integration lengths (see Section 4.2). These observations for $\chi_l = 400\,{\rm Mpc}$ and $z = 0.23$ are shown, for illustrative purposes, in the upper panel of Figure 4.1. In addition, our random data set A can be thought of as consisting of samples from $P(\Delta p_{\rm u}|M_{500}, z; \chi_l)$ at the considered redshifts z, integration lengths χ_l, and true masses M_{500} (see Section 4.3). Our route to obtain $P(p_{\rm c}|M_{500}, z)$ is first to determine $P(p|M_{500}, z; \chi_l)$ as a function of M_{500} as slices in M_{500} of $P(p, M_{500}|z; \chi_l)$, and then, since $= p_{\rm c} + \Delta p_{\rm u}$, to (formally) deconvolve it with $P(\Delta p_{\rm u}|M_{500}, z; \chi_l)$. We describe this procedure, which takes its name from the deconvolution step, in detail in the rest of this section.

We first group our mock measurements of p for $\chi_l = 400\,{\rm Mpc}$ at $z = 0.23$ into 90 subsets by binning their corresponding values of M_{500} into 90 equally-spaced bins between $2 \times 10^{14} M_\odot$ and $10^{15} M_\odot$. For $z = 0.52$ we apply a similar binning but then combine the 33 bins of the high-mass end into groups of three, yielding a total of 68 bins. We do this in order to compensate for the very small number of clusters per original bin at the high-mass end for $z = 0.52$, which in some cases are below 10. We then assume that the distribution followed by p in each of these bins is $P(p|M_{500}, z; \chi_l = 400\,{\rm Mpc})$, with M_{500} equal to the central value of the bin. This is certainly true in the limit in which the bins are infinitesimal, and should be a good approximation if they are sufficiently small. For illustrative purposes, an estimate of $P(p|M_{500}, z; \chi_l = 400\,{\rm Mpc})$ as a function of M_{500} can be seen in the lower panel of Figure 4.1 for $z = 0.23$ (the same case as the upper panel). It was produced with a kernel density estimation method, fastKDE (O'Brien et al., 2014, 2016); we note that this illustrative estimate is done with no previous binning.

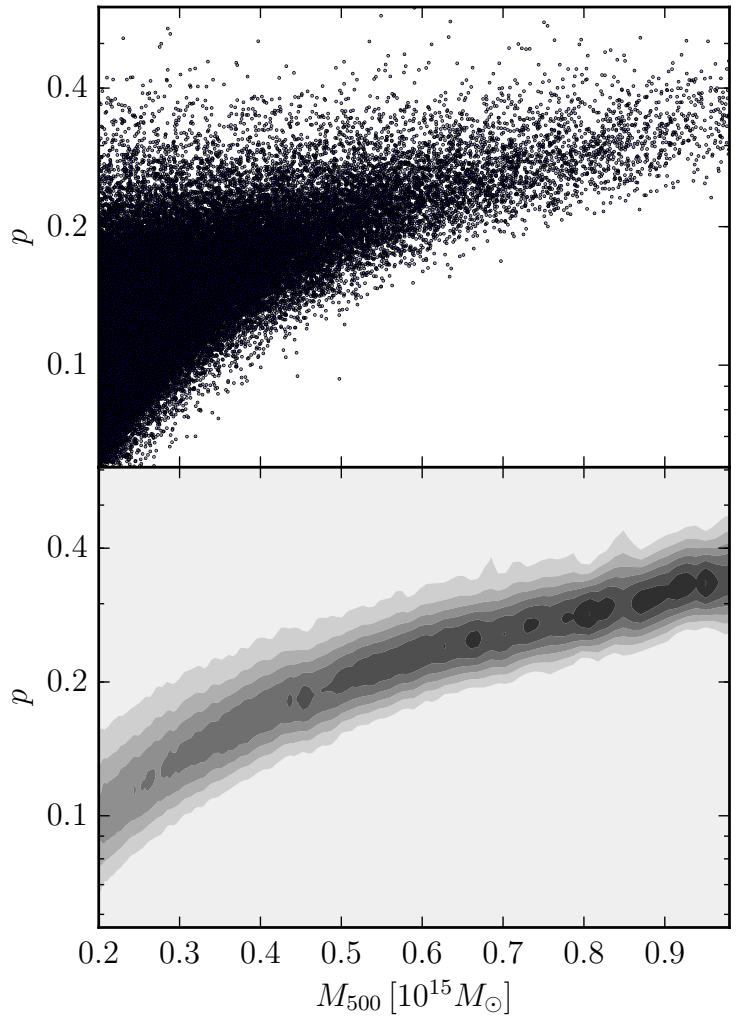


Fig. 4.1 Upper panel: scatter plot of (p, M_{500}) for $\chi_l = 400\,\mathrm{Mpc}$ and $z = 0.23$. Lower panel: estimate of the conditional pdf of p, $P(p|M_{500}, z; \chi_l)$, as a function of M_{500} for the same case as the upper panel, as obtained with the kernel density estimation package fastKDE.

We treat each of the M_{500} bins independently. Let us consider one of our snapshot redshifts z and one of our M_{500} bins. We can write $P(p|M_{500}, z; \chi_l = 400\,\mathrm{Mpc})$ as

$$P(p|M_{500}, z; \chi_l = 400\,\mathrm{Mpc}) = \int_0^\infty P(p|p_\mathrm{c}, M_{500}, z; \chi_l = 400\,\mathrm{Mpc}) P(p_\mathrm{c}|M_{500}, z) dp_\mathrm{c}\,, \quad (4.15)$$

where $P(p|p_\mathrm{c}, M_{500}, z; \chi_l = 400\,\mathrm{Mpc})$ is equal to $P(\Delta p_\mathrm{u}|M_{500}, z; \chi_l = 400\,\mathrm{Mpc})$ evaluated at $\Delta p_\mathrm{u} = p - p_\mathrm{c}$, and where $P(p_\mathrm{c}|M_{500}, z)$ is the distribution of intrinsic scatter that we want to characterise (note that we have dropped the dependence on χ_l following the convention introduced above, as this pdf for $\chi_l = 400\,\mathrm{Mpc}$ ought to be very close to that for $\chi_l = \chi_\star$). That is, as anticipated, the pdf of p can be seen as a convolution of the pdf of p_c with the pdf of Δp_u. We then assume that $P(p_\mathrm{c}|M_{500}, z)$ is approximately log-normal. Log-normality is a common assumption for the intrinsic scatter of cluster observables, even if evidence for, e.g., some skewness has been found in some studies (see, e.g., Becker and Kravtsov 2011). In particular, we model $P(\ln p_\mathrm{c}|M_{500}, z)$[3] with an Edgeworth series, truncating it after the first three expansion terms. That is, we assume that $P(\ln p_\mathrm{c}|M_{500}, z)$ can be written as (Wallace, 1958)

$$P(\ln p_\mathrm{c}|M_{500}, z) = \frac{1}{\sigma_\mathrm{c}} \left(\phi - \frac{\lambda_{3,\mathrm{c}}}{6}\phi^{(3)} + \frac{\lambda_{4,\mathrm{c}}}{24}\phi^{(4)} + \frac{\lambda_{3,\mathrm{c}}^2}{72}\phi^{(6)} \right) . \quad (4.16)$$

Here, ϕ is the standard normal distribution evaluated for $(\ln p_\mathrm{c} - \mu_\mathrm{c})/\sigma_\mathrm{c}$, and $\phi^{(n)}$ is the n-th derivative of the standard normal distribution, also evaluated for $(\ln p_\mathrm{c} - \mu_\mathrm{c})/\sigma_\mathrm{c}$. The quantities μ_c, σ_c, $\lambda_{3,\mathrm{c}}$, and $\lambda_{4,\mathrm{c}}$ are, respectively, the mean, standard deviation, skewness, and kurtosis of $\ln p_\mathrm{c}$. The skewness and kurtosis, in particular, are defined in terms of the cumulants k_n of $\ln p_\mathrm{c}$ as $\lambda_{3,\mathrm{c}} = k_3/\sigma_\mathrm{c}^3$ and $\lambda_{4,\mathrm{c}} = k_4/\sigma_\mathrm{c}^4$, respectively. As can be seen from Eq. (4.16), a Gaussian distribution has $\lambda_{3,\mathrm{c}} = \lambda_{4,\mathrm{c}} = 0$ (and all higher-order cumulants vanish also). We note that $\lambda_{4,\mathrm{c}}$ is sometimes referred to as excess kurtosis. We refer to this model of the intrinsic scatter as the log-Edgeworth model of the intrinsic scatter.

Assuming this model, $P(p_\mathrm{c}|M_{500}, z)$ is characterised by four parameters, which we choose to be β_c, σ_c, $\lambda_{3,\mathrm{c}}$, and $\lambda_{4,\mathrm{c}}$. Here, β_c is a lensing mass bias parameter that is used in substitution of μ_c and is defined as follows. As in Chapter 3, we introduce the

[3]Here and throughout, ln refers to the natural logarithm.

model mean signal-to-noise at true cluster parameters M_{500} and z, $\bar{p}(\beta_c M_{500}, z)$, as

$$\bar{p}\left(\beta_c M_{500}, z\right) \equiv \frac{\kappa_0(\beta_c M_{500}, z)}{\sigma_{\kappa_0}(\beta_c M_{500}, z)} = \frac{\beta_c M_{500}}{\sigma_{M_{500}}\left(\beta_c M_{500}, z\right)}\,, \tag{4.17}$$

where $\kappa_0(\beta_c M_{500}, z)$ is the convergence of the truncated NFW model at the scale radius for a cluster of mass $\beta_c M_{500}$, $\sigma_{\kappa_0}(\beta_c M_{500}, z)$ is the noise for the matched filter in this case, and $\sigma_{M_{500}}\left(\beta_c M_{500}, z\right)$ is the noise of the $\hat{M}_{500}$ matched filter estimator, given by Eq. (2.24). Following Chapter 3, the lensing mass bias parameter β_c is determined by demanding that $\ln \bar{p}(\beta_c M_{500}, z)$ is the mean of $\ln p_c$ for true clusters of mass M_{500} at redshift z, i.e., $\ln \bar{p}(\beta_c M_{500}, z) = \mu_c$. As we discuss further below, deviations of β_c from unity arise both from mismatch between the true mean cluster (plus correlated LSS) convergence and the truncated NFW model at the same mass, and from intrinsic scatter. Although the model mean cluster signal-to-noise is not the expected value of any of our random variables, for small σ_c it is approximately equal to the expected value of p_c, in which case the bias is determined by profile mismatch. We note that in Chapter 3 β_c is denoted with $1 - b_{\rm CMBlens}$ and is assumed to be independent of mass and redshift.

The mean signal-to-noise $\bar{p}$, as defined in Eq. (4.17), is shown in Figure 4.2 for our reference *Planck*-like experiment as a function of M_{500} for the two redshifts and the two matched filter inverse-variance weightings considered, one including the contribution from uncorrelated LSS (dashed lines), and the other one consisting only of reconstruction noise (solid lines). We remind that the former weighting is the one considered in this deconvolution approach; the other choice is also shown for comparison. It can be seen that including LSS in the inverse-variance weighting reduces the mean signal-to-noise at any given mass, as the observations are understood as being noisier. This decrement, however, is small, since reconstruction noise dominates over the variation due to uncorrelated LSS in our reference experiment. We note that $\bar{p}$ in Figure 4.2 is computed assuming $\beta_c = 1$ in Eq. (4.17); $\bar{p}$ can be evaluated at any biased mass $\beta_c M_{500}$ through simple interpolation.

Finally, we use our random data set A to determine the remaining factor of the integrand of Eq. (4.15), $P(\Delta p_u | M_{500}, z; \chi_l = 400\,\mathrm{Mpc})$. Recall, this data set consists of 10^5 mock measurements of Δp_u for each mass bin, redshift, and integration length considered. Specifically, for each of our M_{500} bins we estimate $P(\Delta p_u | M_{500}, z; \chi_l = 400\,\mathrm{Mpc})$ with the corresponding 10^5 mock measurements for $\chi_l = 400\,\mathrm{Mpc}$ in a non-parametric way using the kernel density estimation library fastKDE (O'Brien et al., 2014, 2016).

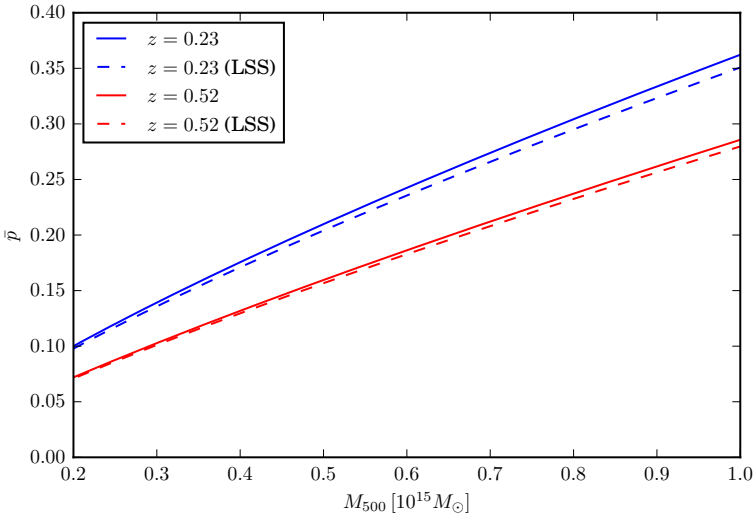


Fig. 4.2 Model mean signal-to-noise, $\bar{p}\,(\beta_{\rm c} M_{500}, z)$, as defined in Eq. (4.17), as a function of M_{500} for $z = 0.23$ (blue curves) and $z = 0.52$ (red curves) and for our two choices of matched filter inverse-variance weighting: one including only the reconstruction noise power spectrum (solid curves); and the other one also including the lensing convergence power spectrum (dashed curves). All the curves are computed using Eq. (4.17) assuming $\beta_{\rm c} = 1$.

Assuming our log-Edgeworth model for the intrinsic scatter and with the scatter due to uncorrelated LSS determined with kernel density estimation, Eq. (4.15) can be thought of as defining the likelihood of our mock measurements of p for $\chi_l = 400\,{\rm Mpc}$ for the four parameters, $\beta_{\rm c}$, $\sigma_{\rm c}$, $\lambda_{3,{\rm c}}$, and $\lambda_{4,{\rm c}}$. Adopting wide flat priors for all four parameters, we explore the corresponding posterior distributions across our M_{500} bins and for our two snapshot redshifts with the `emcee` package[4], generating a total of 10^5 samples for each bin. In addition, we consider a second case in which we take the intrinsic scatter $P(p_{\rm c}|M_{500}, z)$ to be log-normal. This model is the particular case of our log-Edgeworth model in which $\lambda_{3,{\rm c}} = \lambda_{4,{\rm c}} = 0$. In this case, the likelihood has only two parameters, $\beta_{\rm c}$ and $\sigma_{\rm c}$. Similarly adopting wide flat priors on the parameters, we explore the corresponding posteriors with the `emcee` package, also generating 10^5 samples for each bin.

[4]http://dfm.io/emcee/current/

Results and discussion

Figure 4.3 shows the measured values of β_c, σ_c, $\lambda_{3,c}$, and $\lambda_{4,c}$ as a function of M_{500} for our two snapshot redshifts that we obtain with our deconvolution approach. Specifically, for each M_{500} bin the median value of each parameter as obtained from the MCMC samples is shown. The green curves are obtained assuming the log-Edgeworth model of the intrinsic scatter described in Section 4.4.2, whereas the red curves are obtained taking the intrinsic scatter to be log-normal. By construction, $\lambda_{3,c} = \lambda_{4,c} = 0$ for the log-normal case; these constant zero lines are not shown for clarity. The curves are, as expected, noisy, due to the fact that there is only a finite number of clusters in each M_{500} bin. In general, the noise visibly increases with mass, as the number of clusters per bin decreases. For $z = 0.52$, however, the noise decreases at the high-mass end due to the use of wider bins.

In addition, Figure 4.4 shows the histograms of $\ln p$ for our simulated clusters (with $\chi_l = 400\,\mathrm{Mpc}$) in several selected M_{500} bins at both snapshot redshifts. The predictions of our two models, Eq. (4.15) with log-Edgeworth or log-normal intrinsic scatter, evaluated at the respective median parameter values obtained from our MCMC samples, are also shown as the green and red curves, respectively.

Several remarks can be made about Figs 4.3 and 4.4. First, in Figure 4.3 it can be seen that the log-Edgeworth model detects significant non-zero (positive) skewness, $\lambda_{3,c}$, and kurtosis, $\lambda_{4,c}$, across most of our mass bins and for both redshifts. At fixed redshift, $\lambda_{3,c}$ is observed to decrease with mass, with no significant detection at the high mass end. On the other hand, $\lambda_{4,c}$ is observed to depend less strongly on mass and to be non-zero at all masses. Neither of these parameters exhibit a strong redshift dependence. The bias, β_c, and scatter, σ_c, obtained assuming the log-Edgeworth model take values close to those obtained assuming log-normality. A small systematic difference is observed at lower masses, especially for $z = 0.23$, which we attribute to the detection of non-zero skewness and kurtosis, since these two parameters have some degeneracy with β_c and σ_c.

For both models, β_c, is observed to depend on both mass and redshift. At fixed redshift, it decreases with mass. At fixed mass, it is observed consistently to take larger values at $z = 0.52$ than at $z = 0.23$, the difference being larger at smaller masses. The scatter, σ_c, on the other hand, does not have a strong dependence on mass for either model. However, there does seem to be a small dependence with redshift; at a given mass, σ_c generally takes larger values at $z = 0.52$.

We recall that σ_c is the standard deviation of $\ln p_c$, the logarithm of the contribution to the signal-to-noise associated with the cluster itself and to correlated LSS. It is

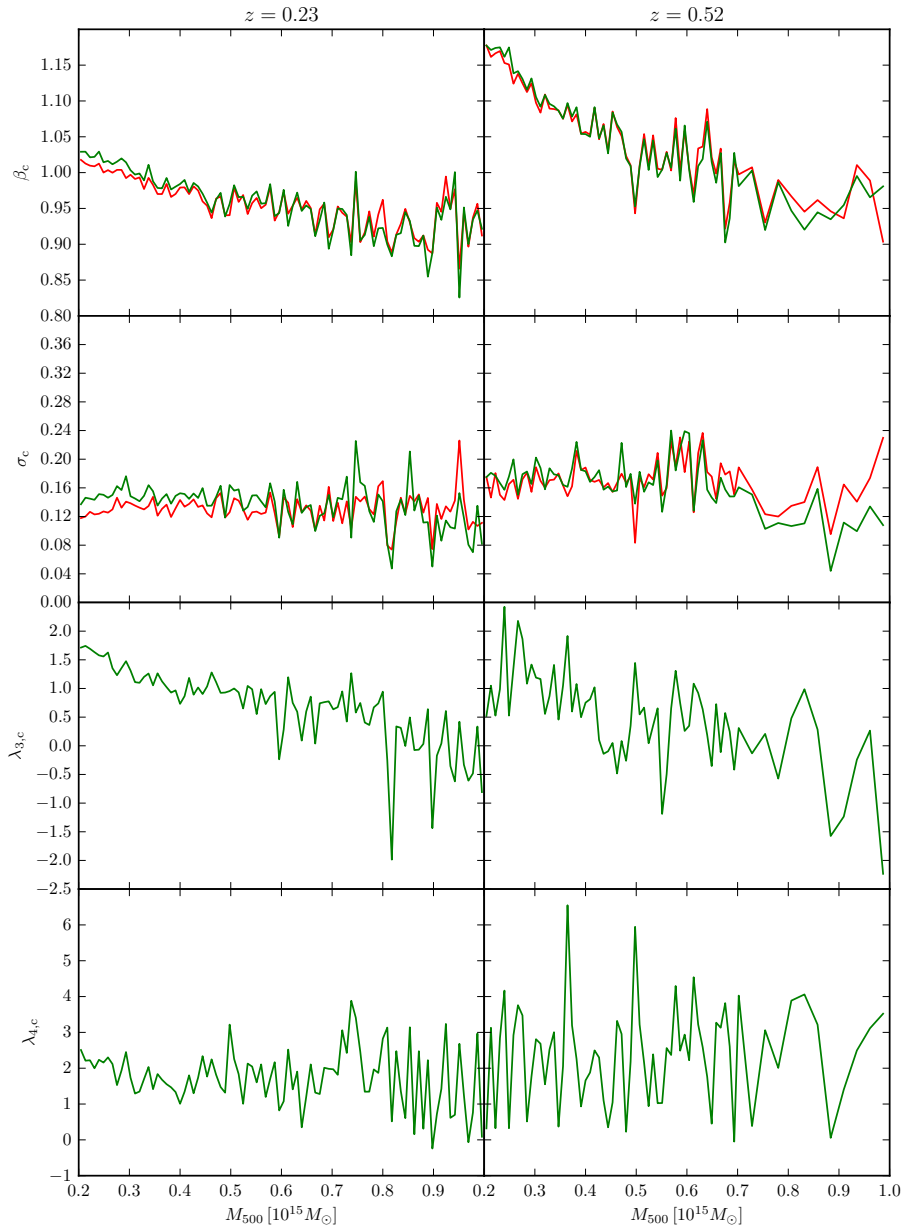


Fig. 4.3 Measured (median) values of β_c (lensing mass bias), σ_c (scatter), $\lambda_{3,c}$ (skewness), and $\lambda_{4,c}$ (kurtosis) as a function of M_{500} for $z = 0.23$ (left) and $z = 0.52$ (right) obtained from the cluster and random data sets A following our deconvolution approach (see Section 4.4.2). The parameter values obtained assuming our log-Edgeworth model are shown in green, whereas those obtained assuming log-normality are shown in red.

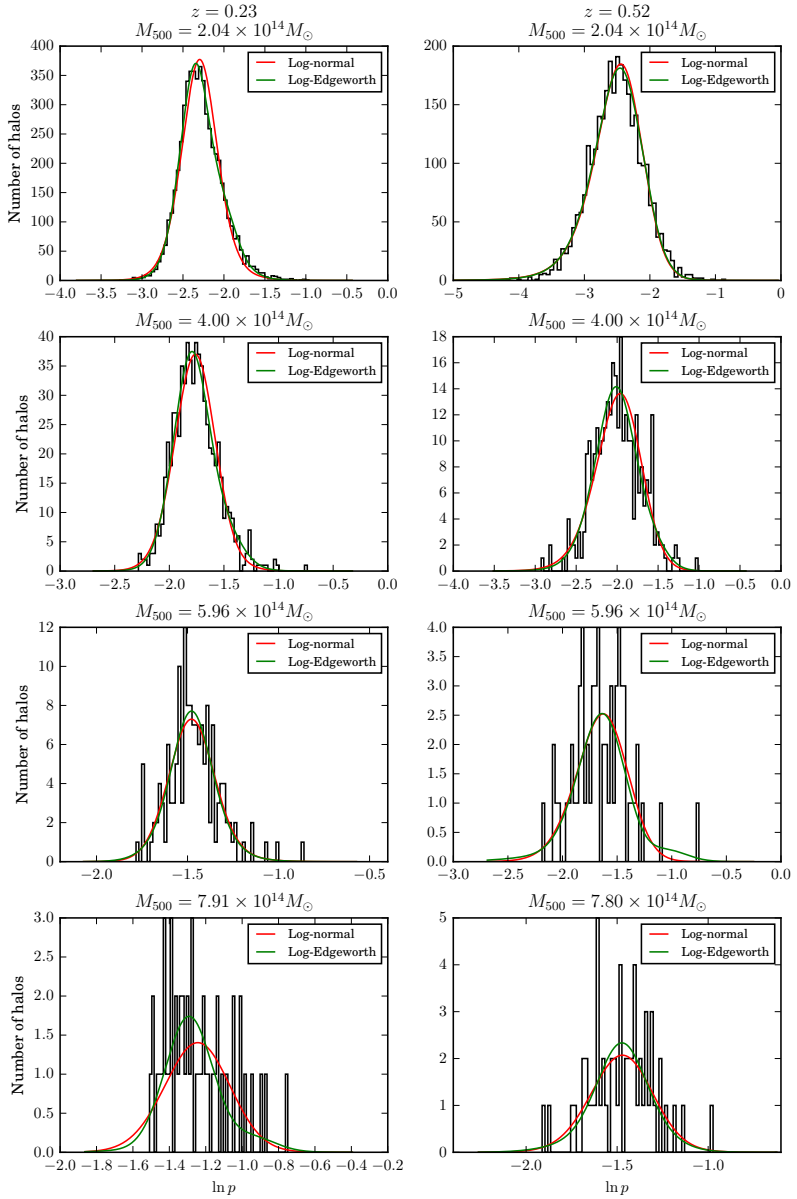


Fig. 4.4 Histograms of the values of $\ln p$ from the cluster data set A for $\chi_l = 400\,\mathrm{Mpc}$ for several selected M_{500} bins and at the two considered snapshot redshifts, $z = 0.23$ (left panels), and $z = 0.52$ (right panels). In addition, corresponding predictions from the log-Edgeworth model of the intrinsic scatter for the median parameter values as obtained from the MCMC samples are shown in green, and analogous predictions from the log-normal model of the intrinsic scatter (see Section 4.4.2) are shown in red.

therefore a measure of how large the fractional scatter on p_c, $\sigma_{p_\mathrm{c}}/p_\mathrm{c}$, is, where σ_{p_c} is the standard deviation of p_c. Thus, we find that the standard deviation of p_c due to the cluster itself and to correlated LSS increases with mass at roughly the same rate as p_c does.

The bias, β_c, on the other hand, has a less straightforward interpretation. It depends both on the mismatch between the convergence model used in the matched filter (in our case, a truncated NFW profile; see Section 4.3) and the true mean convergence at the given M_{500} and z, and on the scatter (and higher moments) at that point in true cluster parameter space. This is a consequence of β_c being defined through the expected value of $\ln p_\mathrm{c}$, μ_c (and not of, e.g., p_c). We remind that $\mu_\mathrm{c} = \ln \bar{p}\,(\beta_\mathrm{c} M_{500}, z)$, where $\bar{p}\,(\beta_\mathrm{c} M_{500}, z)$ is the model mean signal-to-noise defined in Eq. (4.17). In order understand this better, let us assume that the intrinsic scatter is log-normal. In this case,

$$\ln \bar{p}\,(\beta_\mathrm{c}) = \ln \langle p_\mathrm{c} \rangle - \frac{\sigma_\mathrm{c}^2}{2}\,, \tag{4.18}$$

where $\langle p_\mathrm{c} \rangle$ is the expected value of p_c, which only depends on the true mean convergence at the point in cluster parameter space considered, and which is equal to the true expected value of p, since uncorrelated large scale structure has zero expected value. Note that to simplify the notation, we have only made explicit the dependence of $\bar{p}$ on β_c in Eq. (4.18). Thus, in general, at fixed M_{500} and z, β_c depends on both the true mean cluster convergence (through $\langle p_\mathrm{c} \rangle$) and the scatter, σ_c. However, in our case $\sigma_\mathrm{c} \approx 0.15$, so $\sigma_\mathrm{c}^2/2 \approx 0.01$, and $\langle p_\mathrm{c} \rangle$ is typically 0.2, so $\ln \langle p_\mathrm{c} \rangle \sim -1.6$. Therefore, $\ln \bar{p}\,(\beta_\mathrm{c}) \approx \ln \langle p_\mathrm{c} \rangle$, i.e., the bias, β_c, essentially corrects for the mismatch between the assumed convergence model used in the matched filter and the true mean convergence. This is also true if we approximate the intrinsic scatter with our Edgeworth series; indeed, in this case we find

$$\ln \bar{p}\,(\beta_\mathrm{c}) = \ln \langle p_\mathrm{c} \rangle - \frac{\sigma_\mathrm{c}^2}{2} - \ln\left(1 + \frac{\lambda_{3,\mathrm{c}}}{6}\sigma_\mathrm{c}^3 + \frac{\lambda_{4,4}}{24}\sigma_\mathrm{c}^4 + \frac{\lambda^2}{72}\sigma_\mathrm{c}^6\right)\,. \tag{4.19}$$

For $\sigma \approx 0.15$, $\lambda_{3,\mathrm{c}} \approx 1$, and $\lambda_{4,\mathrm{c}} \approx 2$, the new (logarithm) term is $\mathscr{O}(10^{-4})$, and therefore also negligible. This also means that, if the scatter is small, the skewness and kurtosis also have little impact on the bias. We will see in Section 4.4.2 that using a model that better fits the true mean convergence yields $\beta_\mathrm{c} \approx 1$. In addition, in Section 4.4.3, where we consider the extrapolation approach, we will meet an example of the scatter not being negligible and the bias being sensitive to it as a result.

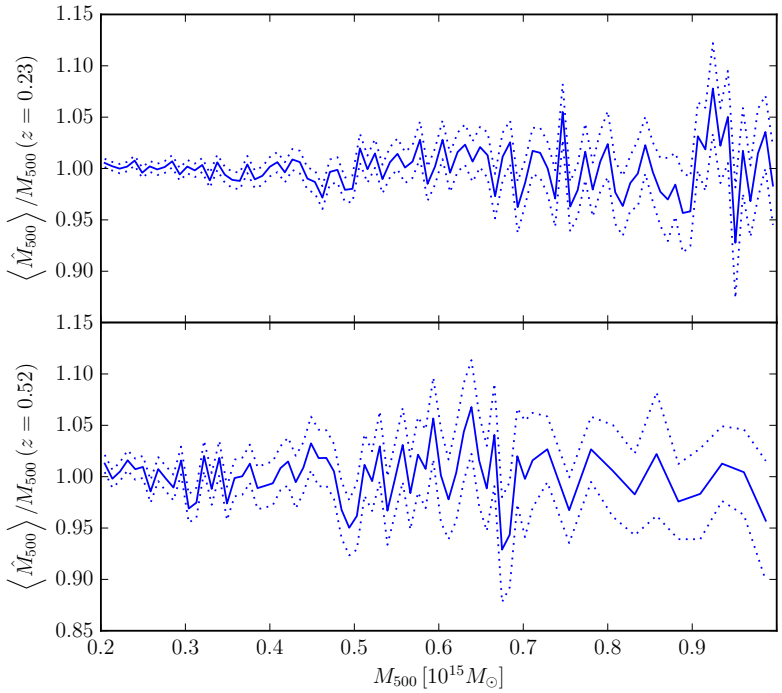


Fig. 4.5 Ratio of the expected value of the estimated mass, $\langle \hat{M}_{500} \rangle$, over the true mass, M_{500}, as a function of M_{500} for $z = 0.23$ (solid curve, upper panel) and $z = 0.52$ (solid curve, lower panel), obtained as detailed in Section 4.4.2. The dotted curves depict the associated standard deviations. No evidence for a bias is seen.

Finally, in Figure 4.4 it can be seen that both models fit our mock data rather well across the mass bins and for both redshifts. Visually, the log-Edgeworth model (green curves) fits the mock data better than the log-normal model (red curves), as expected since the latter is a special case of the former and since statistically-significant skewness and kurtosis are detected throughout the mass bins. This is most apparent in the histograms corresponding to lower masses, which are less noisy as they contain a larger number of clusters (see, e.g., the upper-left panel). We note that rigorous analysis of the goodness of fit of each model and model comparison between the two models, which could be done by, e.g., comparing their Bayesian evidences as a function of mass and redshift, are beyond the scope of this work.

Sufficiency of log-normal approximation

In the previous section we presented evidence for the intrinsic scatter not being log-normal (see Figure 4.3). A question that arises as a consequence is whether assuming

log-normal intrinsic scatter is a good enough approximation. Here we show that for *Planck* clusters log-normal intrinsic scatter is indeed a good approximation that yields unbiased estimates of M_{500}.

We show that assuming log-normal intrinsic scatter is adequate as follows. First, we fit a quadratic polynomial to our median values of β_c and σ_c obtained assuming log-normal intrinsic scatter (red curves in Figure 4.3). We do this by minimising the mean square error, taking the noise associated to each data point (i.e., to each M_{500} bin) to be inversely proportional to the square root of the number of halos within that bin. This produces smooth estimates of β_c and σ_c as a function of M_{500} which, for each M_{500} bin, are much less sensitive to the specific random fluctuation in that bin than the original data point is. Then, for each M_{500} bin we use the likelihood that we used to determine β_c and σ_c, defined in Eq. (4.15), assuming log-normal intrinsic scatter. This function, fed with our smooth estimates of β_c and σ_c (which we evaluate at the central mass of the bin) and with the corresponding mock measurements of p, can be seen as a posterior for M_{500} (after taking the prior on M_{500} to be flat and wide). Since it is a one-dimensional posterior, we explore it by evaluating it for a range of values of M_{500} and we compute the expected value, which we denote with $\langle \hat{M}_{500} \rangle$, and the standard deviation, which we denote with $\sigma_{\hat{M}_{500}}$. Figure 4.5 shows this mean estimated mass over the true mass M_{500}, $\langle \hat{M}_{500} \rangle / M_{500}$, as a function M_{500} for our two snapshot redshifts, along with the corresponding standard deviation, $\sigma_{\hat{M}_{500}}/M_{500}$ (depicted with the dotted lines). No evidence for a bias is seen.

Impact of the choice of cluster convergence model

As argued in Section 4.4.2, at a given point in true cluster parameter (M_{500}–z) space, the mismatch between the lensing convergence model used in the matched filter and the true mean lensing convergence profile is the main contribution to the bias, β_c. In particular, if the model matches the true mean lensing profile, then $\ln \bar{p} = \ln \langle p_c \rangle$ and therefore $\beta_c \approx 1$. In principle, any reasonable model can be used, and any mismatch between it and the true mean convergence profile will make $\beta_c \neq 1$; its precise value can be determined as we have described in Section 4.4.2. However, in a matched filter the signal-to-noise is optimal when the model matches exactly the true signal to be filtered; therefore, using a different template may yield suboptimal results. In this section we validate numerically that $\beta_c \approx 1$ when the model fits the true mean convergence well enough by studying a new set of mock match-filtered observations produced with a more realistic convergence model and we investigate this potential issue of suboptimality.

The model that we use in the matched filter in order to produce the mock measurements studied in this chapter is, as described in Section 4.3, an NFW profile with fixed concentration $c_{500} = 3$ (critical) and truncated at $R_{\rm trunc} = 5R_{500}$. This was also the model used in Chapter 3. Here we consider a more realistic cluster convergence model, which we write as

$$\kappa_{\rm m}(M_{500}, z; b, x_{\rm trunc}) = \kappa_{\rm 1h}(M_{500}, z; b, x_{\rm trunc}) + \kappa_{\rm 2h}(M_{500}, z)\,. \tag{4.20}$$

Here, $\kappa_{\rm 1h}(M_{500}, z; b, x_{\rm trunc})$ is the convergence of a truncated NFW profile with a concentration $c_{500} = 2.6$, which is closer to the values reported by Diemer and Kravtsov (2015) for the mass and redshift range considered here (see the left panel of their Figure 2), evaluated at a biased mass bM_{500} and truncated at $R_{\rm trunc} = x_{\rm trunc}R_{500}$. The second term in the model convergence, $\kappa_{\rm 2h}(M_{500}, z)$, is the two-halo term, which accounts for LSS correlated with the halo. We compute it following Oguri and Hamana (2011) and using the Tinker et al. (2010) model of the halo bias. We refer to the model $\kappa_{\rm m}$ as our fitted truncated NFW+2h model; for comparison purposes, in this section we refer to our original model as our truncated NFW model. For each M_{500} bin, we determine b and $x_{\rm trunc}$ by fitting $\kappa_{\rm m}(M_{500}, z; b, x_{\rm trunc})$ to the corresponding mean convergence as obtained from our cluster convergence fields. More specifically, for each M_{500} bin we obtain a mean convergence map by averaging over the convergence maps of all the clusters within it, and then we azimuthally average this mean convergence map over 40 annuli linearly spaced between the halo centre and an angular radius of 90 arcmin. We then fit for b and $x_{\rm trunc}$ by minimising the mean square error between this binned convergence and the prediction from our model, $\kappa_{\rm m}(M_{500}, z; b, x_{\rm trunc})$, taking the variance of each annulus to be inversely proportional to its area.

The azimuthally-averaged lensing convergence profile as a function of angular separation from the halo centre measured from our cluster convergence fields can be seen in Figure 4.6 for three selected M_{500} bins (solid curves) at our two snapshot redshifts, along with the corresponding predictions from our truncated NFW model (dotted curves) and our fitted truncated NFW+2h model (dash-dot curves). Visually, the latter model clearly fits the measured average convergence better. The fitted values of the two free parameters of this model, b and $x_{\rm trunc}$, are shown in Figure 4.7. The bias, b, deviates from $b = 1$ (no bias) at the few-percent level, whereas the truncation parameter, $x_{\rm trunc}$, takes values around 2, indicating that the two-halo term is already significant at $r \sim 2R_{500}$. We recall that in our truncated NFW model we take $x_{\rm trunc} = 5$, which means that, through the use of a larger truncation radius, this model partially accounts for the two-halo term contribution within $r = 5R_{500}$.

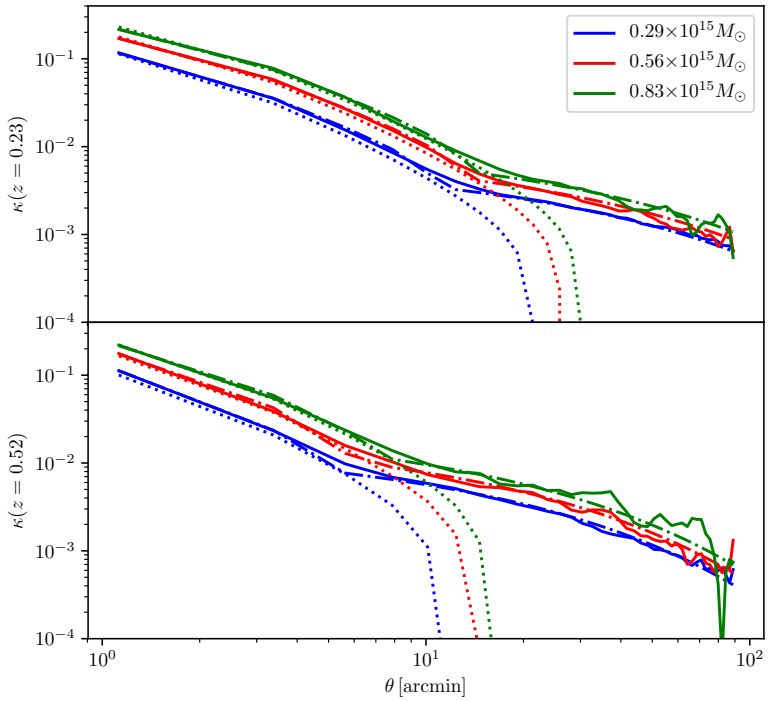


Fig. 4.6 Azimuthally-averaged convergence as a function of angular separation from the cluster centre, θ, for three representative M_{500} bins, each identified by a different colour, for $z = 0.23$ (upper panel) and $z = 0.52$ (lower panel). The solid curves show the empirical convergence as obtained from our convergence maps extracted from the simulation snapshots; the dotted curves show the prediction from our truncated NFW model; and the dash-dotted curves show the prediction from our fitted truncated NFW+2h model. The latter model visibly matches the mean simulation convergence better than the former one.

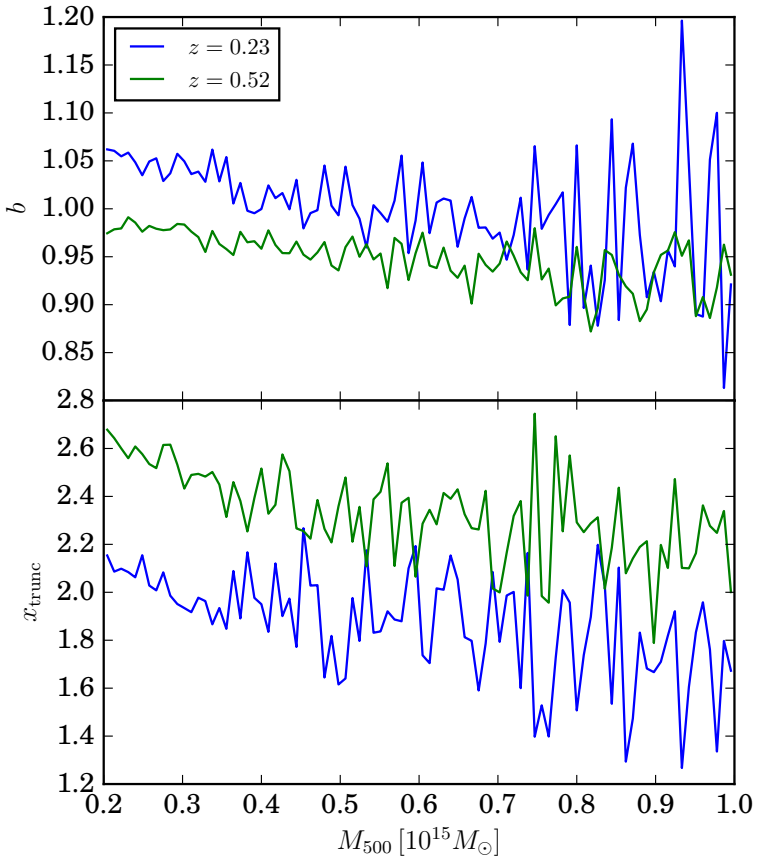


Fig. 4.7 Values of the two free parameters of the fitted truncated NFW+2h model, b and $x_{\rm trunc}$, when fitted to the mean cluster convergence, as a function of M_{500} for $z = 0.23$ (blue curves) and $z = 0.52$ (green curves).

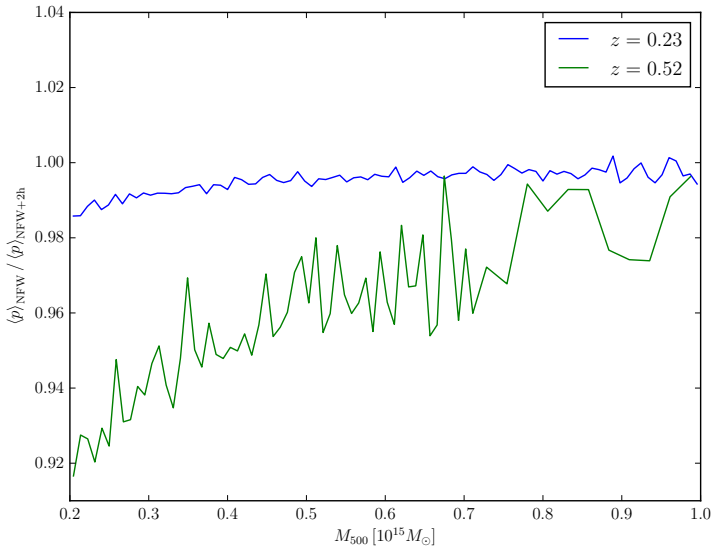


Fig. 4.8 Ratio between the expected value of $p_{\rm c}$, $\langle p_{\rm c} \rangle$, if the truncated NFW convergence model is used in the matched filter and the analogous quantity if the fitted truncated NFW+2h convergence model is used instead, as a function of M_{500} for $z = 0.23$ (blue curves) and $z = 0.52$ (green curves), as obtained from our mock observations.

We then produce a new set of mock observations from our cluster and random convergence maps for $\chi_l = 400\,{\rm Mpc}$ in exactly the same way that we produced our cluster and random data sets used in Section 4.4.2 (i.e., as described in Section 4.3) but using our fitted truncated NFW+2h cluster convergence model in the matched filter instead of our truncated NFW model. The ratio between the expected value of $p_{\rm c}$, $\langle p_{\rm c} \rangle$ (which is equal to the expected value of p, since the contribution from uncorrelated LSS has zero expected value), for this fitted truncated NFW+2h model and the analogous quantity for our truncated NFW model is shown in Figure 4.8 as a function of M_{500} for our two snapshot redshifts. As expected, the fitted truncated NFW+2h model yields a higher mean signal-to-noise, since it fits the signal to be matched by the matched filter better; the difference is larger for $z = 0.52$. Next, we apply our deconvolution approach, as described in 4.4.2, to these new mock observations in order to obtain the values of $\beta_{\rm c}$ and $\sigma_{\rm c}$ for this convergence model; we only consider log-normal scatter. These values are shown in Figure 4.9 as a function of M_{500} for our two snapshot redshifts. As anticipated in Section 4.4.2, $\beta_{\rm c} \approx 1$ throughout our mass bins. At high masses, $\sigma_{\rm c}$ takes similar values to those that we obtain with our truncated NFW model (see Figure 4.3), but at low masses it takes visibly higher values, especially for $z = 0.52$. In fact, for this redshift, at low masses $\beta_{\rm c}$ is visibly smaller than 1, which we attribute

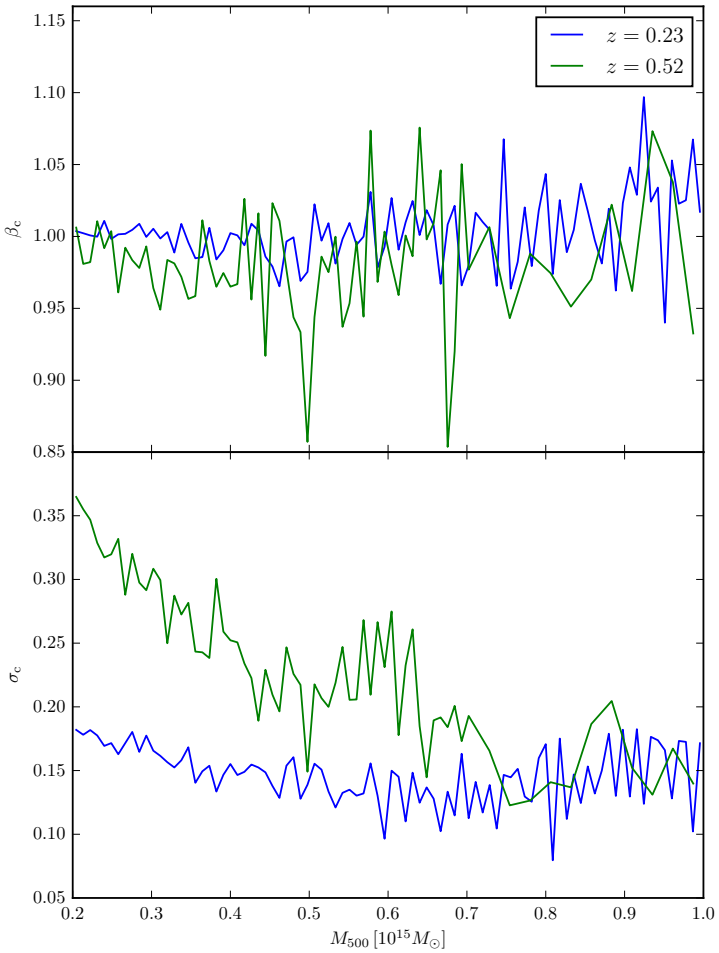


Fig. 4.9 Lensing mass bias, β_c (upper), and scatter, σ_c (lower), as a function of M_{500} for $z = 0.23$ (blue curves) and $z = 0.52$ (green curves) obtained following the deconvolution approach using our fitted truncated NFW+2h convergence model in the matched filter.

to the corresponding larger values of σ_c (as detailed in Section 4.4.2, at fixed mean convergence, a larger value of σ_c yields a lower value of β_c). We attribute this to the fact that it is more spatially extended, and therefore more sensitive to variations around the mean convergence due to correlated LSS.

The fact that the mean signal-to-noise is larger for the fitted truncated NFW+2h model than it is for the truncated NFW model, but that the intrinsic scatter is also larger for the fitted truncated NFW+2h model at lower masses, poses the question of which model is more optimal, i.e., which one yields, after calibration, tighter constraints on M_{500}? We investigate this question by using new mock observations drawn as samples from our hierarchical model. Let us consider one of our two convergence models. First, for each redshift we fit a polynomial to the corresponding measurements of β_c and σ_c as a function of M_{500} obtained assuming log-normality (the posterior median values; i.e., the values shown in Figs 4.3 and 4.9). This yields a smooth estimate of β_c and σ_c as a function of M_{500} for each redshift, which we take to be the true values of these parameters. This smoothing of β_c and σ_c is necessary for the subsequent posteriors of M_{500} to be smooth (see below). Then, using these smooth estimates and Eq. (4.17), we compute $\bar{p}$ at the central value of each of our M_{500} bins. Next, for each bin we draw 10^4 samples from a Gaussian centred at $\ln \bar{p}$ with standard deviation equal to the smooth estimate of σ_c at the bin centre. We exponentiate these samples, which yields a set of samples of p_c for each bin, and we finally add, to each of them, a sample from a Gaussian with zero mean and standard deviation unity. This procedure generates a set of mock observations of p_{obs} drawn according to our hierarchical model (with log-normal intrinsic scatter). We apply this procedure to both convergence models, using the same random seed for each model so that the random fluctuations are the same.

Next, we use these mock observations of p_{obs} for each convergence model in order to constrain, for each mass bin, its corresponding value of M_{500}. We do this by considering the likelihood associated with the hierarchical model used to generate these mock observations: fixing the values of β_c and σ_c to their (smooth) input values, this likelihood can be seen as a function of M_{500}. Assuming a wide flat prior, it yields a posterior for M_{500} for each mass bin. This is very similar to the way in which we obtained a posterior for M_{500} in Section 4.4.2, where we tested the impact of assuming log-normal intrinsic scatter; the only difference is that the pdf with which we convolve the intrinsic scatter pdf is the unit standard-deviation Gaussian that describes the full observational scatter, and not the pdf associated with the uncorrelated LSS within a given integration length, as is the case in Section 4.4.2. Each posterior is one-dimensional, so we explore them

by evaluating them at a range of values of M_{500} and then we compute their standard deviations, which we denote with $\sigma^{\rm NFW}_{\hat{M}_{500}}$ for the truncated NFW convergence model and with $\sigma^{\rm NFW+2h}_{\hat{M}_{500}}$ for the fitted truncated NFW+2h convergence model and which we use as a metric of constraining power.

We find that at a given point in true cluster parameter (M_{500}–z) space, $\sigma^{\rm NFW}_{\hat{M}_{500}}$ and $\sigma^{\rm NFW+2h}_{\hat{M}_{500}}$ are within a few percent of each other. The exact value of their ratio as a function of M_{500} is quite sensitive to the details of the polynomial fit to $\beta_{\rm c}$ and $\sigma_{\rm c}$ (specifically, the polynomial degree and the weighting of each bin), especially at the high-mass end. However, we consistently find some trends. For $z = 0.23$, $\sigma^{\rm NFW+2h}_{\hat{M}_{500}}$ is about 1–2 % larger than $\sigma^{\rm NFW}_{\hat{M}_{500}}$ at low masses, and about 0.5–1 % smaller at intermediate and high masses. That is, the original truncated NFW model slightly outperforms the fitted truncated NFW+2h model at low masses, but becomes less optimal at higher masses. We argue that this is due to the intrinsic scatter $\sigma_{\rm c}$ being larger at low masses for the fitted truncated NFW+2h model (see Figs 4.3 and 4.9). At higher masses, however, the scatters take similar values, and therefore the small difference in mean signal-to-noise (see Figure 4.8), which favours the fitted truncated NFW+2h model, becomes the main source of difference in constraining power. For $z = 0.52$, the results are more sensitive to the fitting details, since there are fewer clusters, especially at the high-mass end, where we cannot determine any trend conclusively. At low masses, however, we find $\sigma^{\rm NFW+2h}_{\hat{M}_{500}}$ to be about 5–10 % larger than $\sigma^{\rm NFW}_{\hat{M}_{500}}$, and about 1–3 % at intermediate masses. As at $z = 0.23$, we attribute this to the (now much larger) intrinsic scatter difference between the two models (see Figs 4.3 and 4.9).

In summary, we find that, at low masses, the fitted truncated NFW+2h convergence model, despite yielding higher signal-to-noise measurements, performs slightly worse than the truncated NFW convergence model due to it having a larger associated intrinsic scatter. For $z = 0.23$, the situation is reversed at intermediate and high masses, where the intrinsic scatter is similar. These differences, however, are small, at the few percent level.

Priors in Chapter 3

As already noted in Section 4.4.2, the approach to understanding the observations considered in this section underlies the hierarchical model of Chapter 3, where $1 - b_{\rm CMBlens}$, $\sigma_{\rm CMBlens}$, and the variable $p_{\rm t}$ correspond to $\beta_{\rm c}$, $\sigma_{\rm c}$, and $p_{\rm c}$ here, respectively. We note, however, that $1 - b_{\rm CMBlens}$ in Chapter 3 is also intended to account for the few-percent negative bias due to cluster miscentering. In that work, the intrinsic scatter is assumed to be log-normal, which, as shown in Section 4.4.2, is a sufficiently good

approximation, and priors are imposed on $1 - b_{\rm CMBlens}$ and $\sigma_{\rm c}$. These are Gaussian priors centred at $1 - b_{\rm CMBlens} = 0.93$ and $\sigma_{\rm CMBlens} = 0.2$, both with standard deviation of 0.05.

The cluster sample used in Chapter 3 is the *Planck* MMF3 cosmology sample, which consists of 439 SZ-detected clusters (see Planck 2015 results XXVII 2016). In it, most clusters are at low redshift (see Figure 1 of Planck 2015 results XXVII 2016), with 246 out of 433 clusters with measured redshift being at $z < 0.23$ ($z = 0.23$ being the redshift of our low-redshift snapshot), and 311 being at $z < 0.3$. Only 22 clusters are at $z > 0.52$ (i.e., above the redshift of our high-redshift snapshot). Also, in Figure 1 of Planck 2015 results XXVII (2016) it can be seen that most cluster masses take values lying around the centre of our considered mass range at lower redshifts, and values towards our high-mass end at higher redshifts. The latter is particularly true for clusters with redshifts close to our high-redshift snapshot. Therefore, most of the clusters in the sample lie within the region of mass–redshift parameter space where the priors imposed on $1 - b_{\rm CMBlens}$ and $\sigma_{\rm CMBlens}$ in Chapter 3 are consistent with our measurements of $\beta_{\rm c}$ (plus a few percent decrement due to miscentering; see Chapter 2) and $\sigma_{\rm c}$, respectively.

4.4.3 Extrapolation approach

Statistics as a function of integration length

In the extrapolation approach, only the reconstruction noise power spectrum is present in the matched-filter inverse-variance weighting (see Eq. 4.8). As in Section 4.4.3, let us first think of a hypothetical set of simulated observations with reconstruction noise and with LSS present from $z = 0$ back to last-scattering. Since reconstruction noise is approximately Gaussian, as a consequence of our choice of matched-filter weighting $P(p_{\rm obs}|p, M_{500}, z; \chi_\star)$ is approximately a unit-variance Gaussian centred on p. We remind that $p = p_{\rm c} + \Delta p_{\rm u}$, and that $\chi_\star$ denotes that integration is performed from $z = 0$ back to last-scattering. In order to fully determine the statistics of $p_{\rm obs}$ at given true cluster parameters, M_{500} and z, what remains to be characterised is $P(p|M_{500}, z; \chi_\star)$. In the context of this approach we refer to the variability due to reconstruction noise, $P(p_{\rm obs}|p, M_{500}, z; \chi_\star)$, as *observational* scatter, and to the remaining variability, due to the cluster itself, correlated and uncorrelated LSS, $P(p|M_{500}, z; \chi_\star)$, as *intrinsic* scatter. We note that this way of understanding the scatter is different from that made in Section 4.4.2, in which the variation due to uncorrelated LSS is understood as being part of the observational scatter.

The aim of the rest of this section is to determine $P(p|M_{500}, z; \chi_\star)$, which we do with our mock observations, namely with our cluster and random data sets B. We first study the statistics of p as a function of M_{500}, z, and integration length, χ_l. We can do this with the mock observations of cluster data set B, which can be thought of as samples from $P(p, M_{500}|z; \chi_l)$, where χ_l can be any of the seven integration lengths that we consider; $P(p|M_{500}, z; \chi_l)$ can be obtained as slices in M_{500} of $P(p, M_{500}|z; \chi_l)$. In particular, for each of our two snapshot redshifts and each of our seven values of χ_l, we divide our data points into 90 subsets, binning M_{500} in 90 equally-spaced bins between $2 \times 10^{14} M_\odot$ and $10^{15} M_\odot$. As in Section 4.4.2, we think of the value of p of each data point falling within a given bin as a sample of $P(p|M_{500}, z; \chi_l)$, M_{500} being the mass at the bin centre.

In order to characterise the distribution $P(p|M_{500}, z; \chi_l)$, we first compute its empirical mean, which we denote with $\langle p \rangle$, with the values of p within each mass bin. We also compute the empirical mean, standard deviation, skewness, and kurtosis of $P(\ln p|M_{500}, z; \chi_l)$, which we denote with μ, σ, λ_3, and λ_4, respectively, as a function of M_{500} and χ_l for our two snapshots. Analogously to what is done in Section 4.4.2, λ_3 and λ_4 are defined in terms of the cumulants of $P(\ln p|M_{500}, z; \chi_l)$, k_n, as $\lambda_3 = k_3/\sigma^3$ and $\lambda_4 = k_4/\sigma^4$, respectively. Also as in Section 4.4.2, instead of presenting our results in terms of μ, we introduce a bias parameter, β, which is defined in an analogous way to β_c in Section 4.4.2, i.e., by imposing that $\ln \bar{p}(\beta M_{500}, z) = \mu$, where $\bar{p}(\beta M_{500}, z)$ is given by Eq. (4.17), noting that $\sigma_{M_{500}}$ is now the noise of the matched filter with only reconstruction noise in the inverse-variance weighting.

Figure 4.10 shows $\langle p \rangle$ and β as a function of integration length for three selected mass bins for the $z = 0.23$ snapshot. We obtain the error bars with bootstrapping, resampling from our cluster data points (M_{500}, p), with a (M_{500}, p) pair taken as a single data point so that the correlation structure is not lost. As expected, the errors are larger for higher masses, as there are fewer clusters per bin. In addition, our measurements of β, σ, λ_3, and λ_4 are shown in Figure 4.11 as a function of M_{500} and integration length for our two redshifts. For the sake of clarity, only four integration lengths have been included.

Since p is linear in the lensing convergence (see Eq. 4.12), $\langle p \rangle$ is linear in the mean lensing convergence within a given integration length at given true cluster parameters and is insensitive to other moments of the convergence. This is illustrated in the upper panel of Figure 4.10: $\langle p \rangle$ responds to the change in the mean convergence as the integration length is increased starting from its lowest value, increasing as more

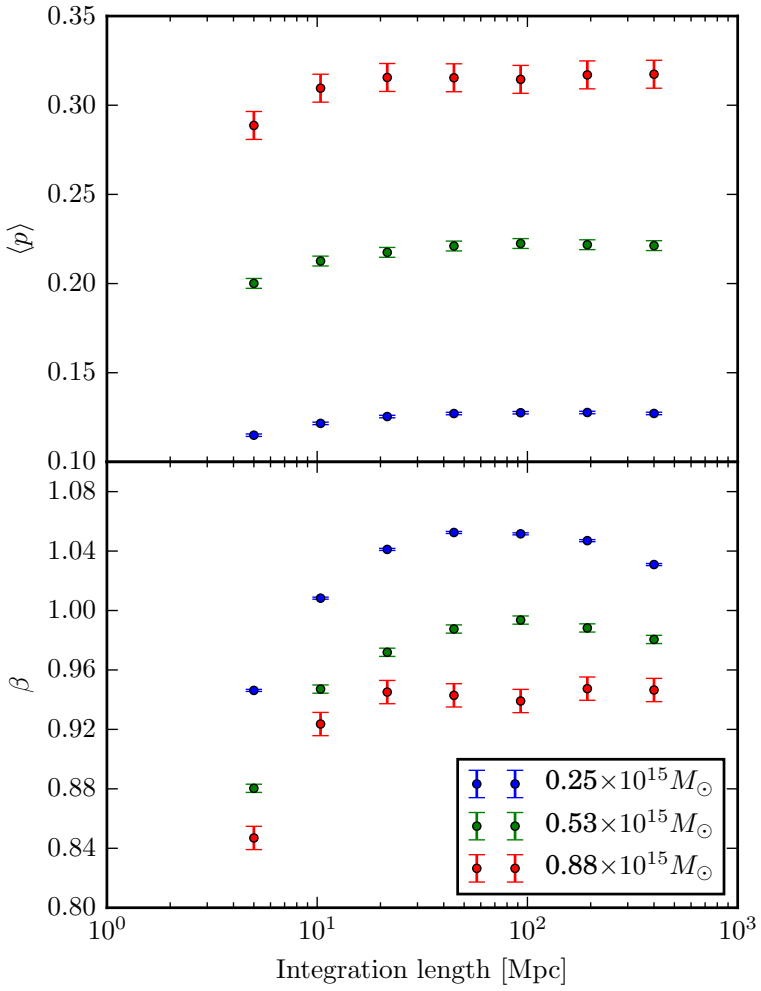


Fig. 4.10 Expected value of p, $\langle p \rangle$, (upper panel) and bias, β, (lower panel) as a function of integration length, χ_l, (in comoving Mpc) for three representative M_{500} bins, as obtained from the cluster data set B (see Section 4.4.3). The error bars are obtained with bootstrapping.

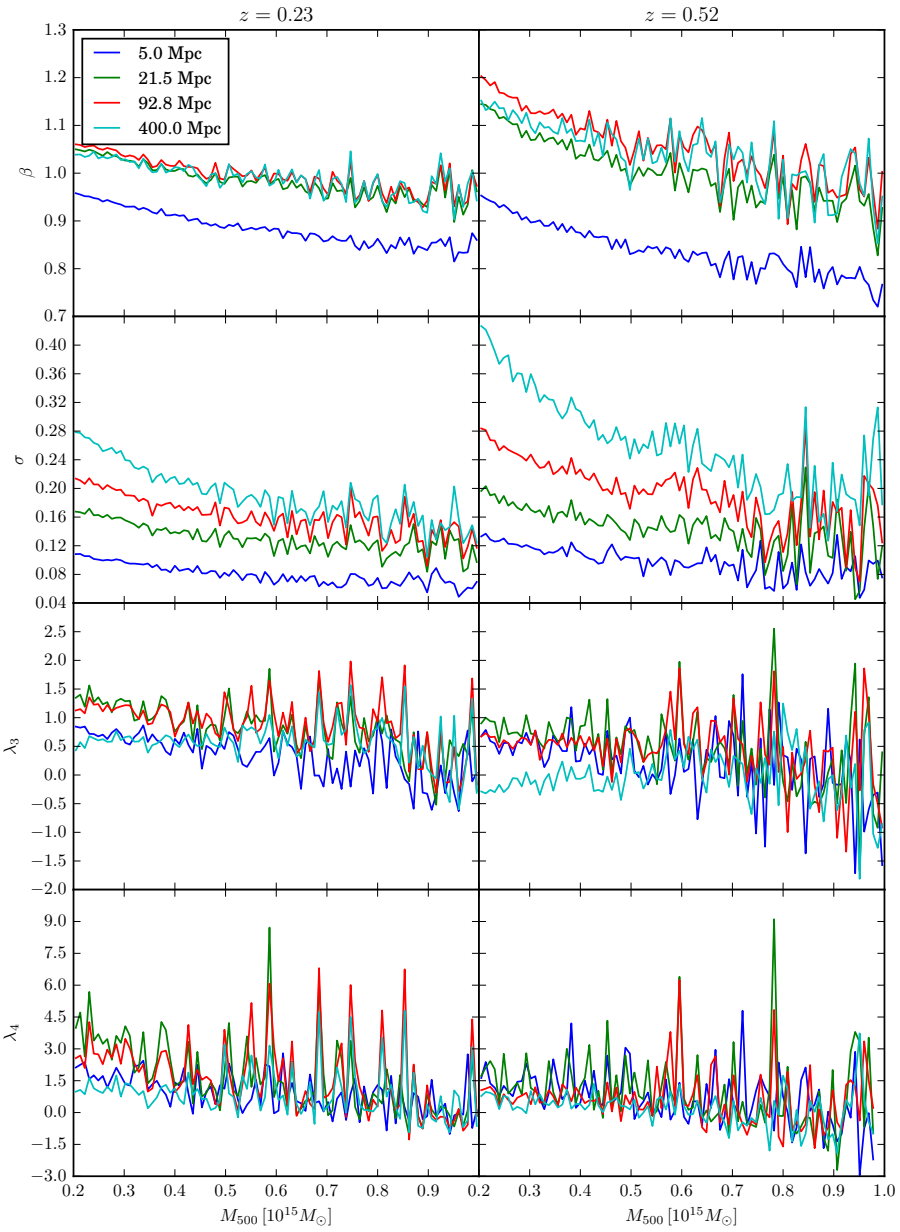


Fig. 4.11 Measured values of β (bias), σ (scatter), λ_3 (skewness), and λ_4 (kurtosis) as a function of M_{500} and integration length, χ_l, for $z = 0.23$ (left) and $z = 0.52$ (right), as obtained from cluster data set B (see Section 4.4.3).

correlated LSS is added, but this growth plateaus at large integration lengths, where effectively only zero-mean uncorrelated LSS is being added.

The bias parameter β, however, defined in terms of $\ln p$, is not linear in the convergence. In general it feels both the effect of the mean convergence at a given integration length and of the variance, and higher moments, around the mean convergence. In Section 4.4.2 we saw that the analogous parameter defined in the context of the deconvolution approach, $\beta_{\rm c}$, is effectively only sensitive to the mean convergence, due to the fact that the intrinsic scatter is small; that is, this analogous parameter essentially corrects the mismatch between the true mean convergence and the model used in the matched filter. Following an argument similar to that in Section 4.4.2, and assuming log-normality (as shown in Eq. 4.19, the effect of the skewness and the kurtosis is higher order in the scatter and therefore negligible),

$$\ln \bar{p}(\beta M_{500}, z) = \ln \langle p \rangle - \frac{\sigma^2}{2} \,. \tag{4.21}$$

At low integration lengths, σ is small, as can be seen in Figure 4.11, and, as in the analogous case in Section 4.4.2, negligible. The bias, β, is therefore just sensitive to the mean convergence profile through $\ln \langle p \rangle$, which at low integration lengths changes with integration length, as there is still correlated LSS that is being added. In particular, we can see in Figure 4.10 that β initially increases with integration length. As the integration length increases, however, the mean convergence changes progressively less, which can be seen in the plateauing of $\langle p \rangle$ in Figure 4.10, and more uncorrelated LSS is added, with σ increasing as a result (see Figure 4.11). Equation (4.21) shows that at fixed $\langle p \rangle$, if σ increases $\bar{p}$ decreases and, as a consequence, so does β. This fall with integration length is observed at high integration lengths in Figs 4.10 and 4.11, being more pronounced at lower masses.

In Figure 4.10 it can also be seen that β and σ decrease with M_{500} for both redshifts, analogously to $\beta_{\rm c}$ and $\sigma_{\rm c}$ in Section 4.4.2. In addition, the skewness, λ_3, and kurtosis, λ_4, exhibit a similar qualitative behaviour as a function of integration length across mass and for both redshifts: they both take lower values for $\chi_l = 5\,{\rm Mpc}$, the shortest integration length considered, grow with χ_l to reach some maximum value, and then decrease beyond this as effectively only uncorrelated LSS is added. That is, above some value of χ_l, as integration length increases the intrinsic scatter, due to the cluster itself, correlated and uncorrelated LSS, becomes increasingly log-normal. As in Section 4.4.2, the skewness and the kurtosis are generally lower for $z = 0.52$ than for $z = 0.23$.

Extrapolation

The fact that for large values of χ_l, at which effectively only uncorrelated LSS is added on increasing χ_l further, $P(p|M_{500}, z; \chi_l)$ becomes increasingly log-normal motivates the following approach to obtain $P(p|M_{500}, z; \chi_\star)$, the ultimate goal of this section.

For a given integration length χ_l, $p = p_\mathrm{c} + \Delta p_\mathrm{u}$. Since p_c and Δp_u are independent random variables, the cumulants of p are equal to the cumulants of p_c plus the cumulants of Δp_u. In particular, the mean of p, $\langle p \rangle$, is equal to the mean of p_c, since Δp_u has zero mean, and $\sigma_p^2 = \sigma_{p_\mathrm{c}}^2 + \sigma_{\Delta p_\mathrm{u}}^2$, where σ_p^2, $\sigma_{p_\mathrm{c}}^2$, and $\sigma_{\Delta p_\mathrm{u}}^2$ are the variances of p, p_c, and Δp_u, respectively.

For $\chi_l = 400\,\mathrm{Mpc}$ we measure $\langle p \rangle$ and σ_p^2 from our cluster data set B and $\sigma_{\Delta p_\mathrm{u}}^2$ from our random data set B across the 90 M_{500} bins. We then obtain $\sigma_{p_\mathrm{c}}^2$ by subtraction. Since this integration length effectively contains all LSS correlated with the cluster, our measured values of $\langle p \rangle$ and $\sigma_{p_\mathrm{c}}^2$ are the same as the values of such cumulants for $\chi_l = \chi_\star$, which we denote with $\langle p \rangle\,(\chi_\star)$ and $\sigma_{p_\mathrm{c}}^2(\chi_\star)$, respectively. Thus, we just need to determine $\sigma_{\Delta p_\mathrm{u}}^2(\chi_\star)$, the variance associated with LSS from $z = 0$ to last-scattering.

In general, $\sigma_{\Delta p_\mathrm{u}}^2(\chi_l)$ can be written as

$$\sigma_{\Delta p_\mathrm{u}}^2(\chi_l) = \left[\int \frac{d^2\boldsymbol{L}}{2\pi} \frac{|\kappa_\mathrm{t}(\boldsymbol{L})|^2}{N_\kappa(\boldsymbol{L})}\right]^{-1} \int \frac{d^2\boldsymbol{L}}{2\pi} \frac{|\kappa_\mathrm{t}(\boldsymbol{L})|^2}{|N_\kappa(\boldsymbol{L})|^2} C_L^{\kappa\kappa}(\chi_l)\,, \tag{4.22}$$

where $C_L^{\kappa\kappa}(\chi_l)$ is the power spectrum of the lensing convergence within the comoving distance interval specified by χ_l, and the other variables are the same as in Eq. (4.12). We compute $C_L^{\kappa\kappa}(\chi_l)$ with CAMB[5] using the nonlinear matter power spectrum and the Limber approximation. We find that for $\chi_l = 400\,\mathrm{Mpc}$ and for $\chi_l = 192.7\,\mathrm{Mpc}$, $\sigma_{\Delta p_\mathrm{u}}^2(\chi_l)$ computed this way is within a few percent of the value we measure across our M_{500} bins and our two redshifts. We also find that, for a given redshift and mass bin, the small fractional disagreement for $\chi_l = 400\,\mathrm{Mpc}$ is very similar to that for $\chi_l = 192.7\,\mathrm{Mpc}$. This motivates us to compute $\sigma_{\Delta p_\mathrm{u}}^2(\chi_\star)$ using Eq. (4.22) with a simple rescaling in each mass bin and at each redshift by the appropriate factor obtained empirically for $\chi_l = 400\,\mathrm{Mpc}$. We add this rescaled value of $\sigma_{\Delta p_\mathrm{u}}^2(\chi_\star)$ to $\sigma_{p_\mathrm{c}}^2(\chi_\star)$ to obtain finally $\sigma_p^2(\chi_\star)$

We then assume that $P(p|M_{500}, z; \chi_\star)$ is log-normal. Denoting the mean and standard deviation of $P(\ln p|M_{500}, z; \chi_\star)$ with μ_e and σ_e, respectively, we compute them by imposing that the mean and variance of $P(p|M_{500}, z; \chi_\star)$ have the values $\langle p \rangle\,(\chi_\star)$ and $\sigma_p^2(\chi_\star)$ that we determine as described above. Following Sections 4.4.2 and

[5] `www.camb.info`

4.4.3, instead of presenting our results in terms of $\mu_{\rm e}$ we introduce a bias parameter, $\beta_{\rm e}$, defined such that $\ln \bar{p}(\beta_{\rm e} M_{500}, z) = \mu_{\rm e}$, with $\bar{p}$ given by Eq. (4.17) (where $\sigma_{M_{500}}$ is now computed with only the reconstruction noise power spectrum in the inverse-variance matched-filter weighting).

The bias, $\beta_{\rm e}$, and scatter, $\sigma_{\rm e}$, obtained this way are shown in Figure 4.12. As expected, the addition of a large amount of uncorrelated LSS significantly decreases $\beta_{\rm e}$ compared to our previous measurements of β for smaller integration lengths (see Figure 4.11 and Eq. 4.21). This is true across our M_{500} range and for both redshifts, the effect being more significant at lower masses. Similarly, the scatter has increased substantially with respect to the values observed for smaller integration lengths (see also Figure 4.11), being dominated by uncorrelated LSS.

We remark that this extrapolation approach relies on the assumption of log-normality of $P(p|M_{500}, z; \chi_\star)$, an assumption that, although motivated by our mock observations (see, e.g., Figure 4.11), we cannot check empirically. A more rigorous approach to study the full line-of-sight intrinsic scatter would involve considering full lightcone simulations of the lensing convergence from $z = 0$ back to last-scattering (see, e.g., Giocoli et al. 2016), something that is beyond the scope of this work.

In summary, this extrapolation approach treats the total scatter due to the cluster and to correlated and uncorrelated LSS as log-normal, and reconstruction noise as Gaussian. In contrast, the deconvolution approach (Section 4.4.2) treats the scatter due to the cluster and to correlated LSS as log-normal (or close to log-normal), and the scatter due to uncorrelated LSS and reconstruction noise as Gaussian. For an experiment like *Planck*, both approaches are essentially equivalent. Indeed, as shown in Section 4.4.2, the scatter due to the cluster and to correlated LSS is approximately log-normal, at least for mid to high-mass clusters, and the scatter due to uncorrelated LSS and reconstruction noise is roughly Gaussian, since reconstruction noise is almost Gaussian and dominates over uncorrelated LSS (which itself is approximately Gaussian on the scales relevant for cluster mass estimation with *Planck*). In addition, as argued in this section, the total scatter due to the cluster and to correlated and uncorrelated LSS is approximately log-normal for a *Planck*-like experiment.

This equivalence may no longer be valid for future experiments, which will have lower reconstruction noise, i.e., higher signal-to-noise per cluster, and probe the convergence on smaller angular scales. If reconstruction noise becomes comparable with the scatter due to uncorrelated LSS, approximating the summed scatter as Gaussian may no longer be accurate. The extrapolation approach, therefore, may be preferable in such cases. We note that the extrapolation approach does not include the LSS convergence

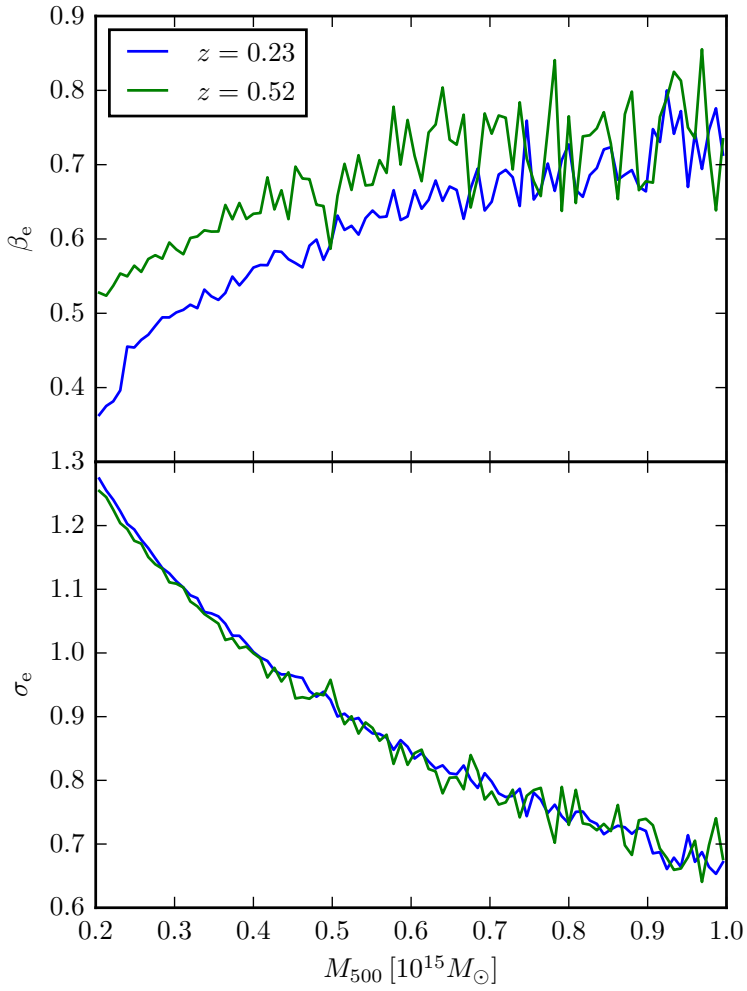


Fig. 4.12 Measured values of the lensing mass bias, β_e (upper panel), and scatter, σ_e (lower panel), extrapolated to include uncorrelated LSS from $z = 0$ to last-scattering for our *Planck*-like reference experiment as a function of M_{500} and for $z = 0.23$ (blue curves) and $z = 0.52$ (green curves).

power spectrum in the inverse-variance weighting of the matched filter, and, as a consequence, does not optimise for it, as opposed to the deconvolution approach. There is therefore some signal-to-noise loss in the extrapolation approach with respect to the deconvolution approach. For *Planck*, which is reconstruction-noise dominated, this loss is small, but it may be more significant in lower reconstruction noise scenarios.

4.5 Different observational specifications

We briefly consider how the results obtained with our extrapolation approach vary if we assume a different experimental setup. In particular, we consider an idealised experiment analogous to the *Planck*-like experiment considered throughout this chapter but with a FWHM $= 1.4\,\mathrm{arcmin}$ and with temperature noise levels of $7\,\mu\mathrm{K}\,\mathrm{arcmin}$; this is intended to resemble AdvACT. We consider two different lensing quadratic estimators, the TT quadratic estimator, which is the one used throughout this chapter for our *Planck*-like experiment, and the EB quadratic estimator. For both quadratic estimators we use a maximum multipole of $l = 2000$ in the gradient leg and of $l = 3000$ in the field leg.

The EB quadratic estimator differs from the TT quadratic estimator in that it takes as input one map of the (lensed) E-mode and one map of the (lensed) B-mode CMB polarisation in order to estimate the lensing convergence, instead of two copies of the (lensed) temperature anisotropies. (See Hu and Okamoto 2002 for the detailed construction of the EB estimator.) The E-modes and B-modes are a useful basis-independent description of the (linear) polarisation of the CMB (Seljak and Zaldarriaga, 1997; Kamionkowski et al., 1997). Lensing reconstruction with polarisation is significantly noisier than reconstruction with temperature for an experiment like *Planck* (Planck 2018 results VIII, 2018). Indeed, for this experiment, for a cluster with $M_{500} = 5 \times 10^{14} M_\odot$ at redshift $z = 0.23$ and with the convergence profile of our matched-filter template (Eq. 4.7), we find, using Eq. (4.17) with $\beta_\mathrm{c} = 1$, a mean signal-to-noise of $\langle p \rangle = 0.21$ if the TT quadratic estimator is used. If the EB quadratic estimator is used instead, we find $\langle p \rangle = 0.03$. The situation is, however, different for our AdvACT-like experiment, for which, for the same cluster, we find $\langle p \rangle = 0.42$ if the TT estimator is used, and $\langle p \rangle = 0.41$ if the EB estimator is used. The EB estimator therefore provides a signal-to-noise similar to that of the TT estimator for this experimental set-up. It also has the additional interest that it is less sensitive to extragalactic foregrounds than the TT estimator. In the cluster context, the TT estimator in particular suffers from contamination due to the thermal and kinetic SZ

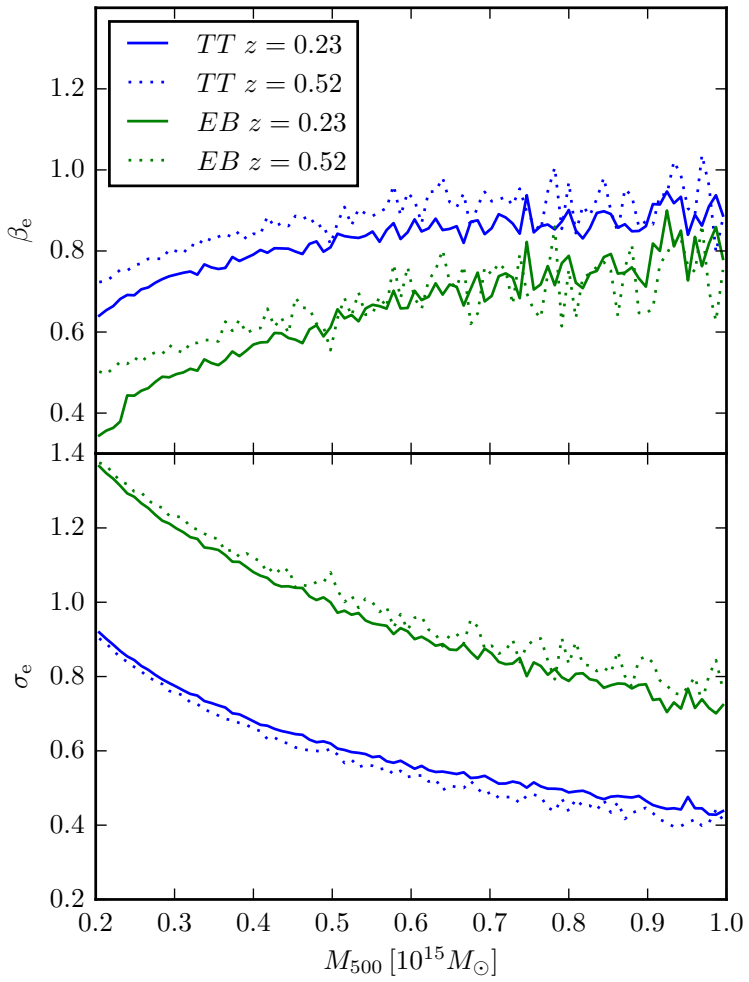


Fig. 4.13 Measured values of the lensing mass bias, β_e (upper panel), and scatter, σ_e (lower panel), extrapolated to include uncorrelated LSS from $z = 0$ to last-scattering, for an AdvACT-like experiment with the TT quadratic estimator (in blue) for $z = 0.23$ (solid curves) and $z = 0.52$ (dotted curves). Analogous results for the EB quadratic estimator are shown in green.

effects from the cluster itself, the latter being difficult to remove given that it has the same frequency dependence as the CMB anisotropies, while the polarisation signal from such effects is a negligible contaminant (Raghunathan et al., 2017). We note that the other three quadratic estimators, the TE, the TB, and the EE estimators, yield, in this scenario, smaller mean signal-to-noise ratios, $\langle p \rangle = 0.22$, $\langle p \rangle = 0.14$, and $\langle p \rangle = 0.24$, respectively. We therefore do not consider them here.

The extrapolated bias, $\beta_{\rm e}$, and scatter, $\sigma_{\rm e}$, that we obtain using the TT and EB quadratic estimators for our AdvACT-like set-up are shown in Figure 4.13. It can be observed that the values of $\beta_{\rm e}$ and $\sigma_{\rm e}$ as a function of M_{500} and z depend on the choice of quadratic estimator and, also, on the experiment specifications (compare with Figure 4.12, where analogous results are shown for the *Planck*-like experiment). This implies that, although the calibration approaches presented in this work may be applied to other lensing estimation techniques and experimental set-ups, their numerical results require case-by-case consideration.

4.6 Conclusion

We have studied the statistics of a CMB lensing galaxy cluster mass observable, p, for a *Planck*-like experiment with mock observations obtained from an N-body simulation, characterising the biased mean signal and the scatter, and deviations from log-normality, due to the variation associated with the cluster and with correlated and uncorrelated LSS. This characterisation is essential for a cosmological analysis that may make use of this mass observable (e.g., Chapter 3) to deliver unbiased results.

We have followed two alternative routes in order to quantify the statistics of this mass observable. First, in our deconvolution approach (Section 4.4.2), we have treated the variation due to uncorrelated LSS as noise in the matched-filtering process, where it adds to the reconstruction noise, and then characterised the mean signal and the variation due to the cluster itself and to correlated LSS (what we call intrinsic scatter) with our mock observations. We find this intrinsic scatter to be roughly log-normal, although significant skewness and kurtosis are detected, as in similar studies of galaxy weak lensing cluster mass measurements (e.g., Becker and Kravtsov 2011, where significant positive skewness is found). This approximate log-normality and our measured values of the bias, $\beta_{\rm c}$, and the scatter, $\sigma_{\rm c}$, serve as justification for the priors on the analogous parameters used in Chapter 3. Second, in our extrapolation approach (Section 4.4.3) we have considered the variation due to the cluster itself and to both correlated and uncorrelated LSS as intrinsic scatter. We find with our mock

observations that this scatter becomes increasingly log-normal as we integrate along longer paths, which motivates us to extrapolate our results in order to incorporate the scatter due to uncorrelated LSS from $z = 0$ back to CMB last-scattering assuming log-normality.

We have also considered, for illustration, how our extrapolation results change if a different experimental set-up is assumed (an AdvACT-like experiment; see Section 4.5). While the qualitative trends with mass and redshift are similar, numerical results for β_e and σ_e differ significantly from those obtained for the *Planck*-like experiment. This implies that ongoing and future experiments that may want to use our CMB lensing mass observable for, e.g., mass calibration in a cluster counts analysis, will have to quantify its statistics for the particular case of their experiment specifications. Our numerical results, for both a *Planck*-like and an AdvACT-like experiment, are not transferable to other experiments. Different cluster mass observables (e.g., the observable proposed in Raghunathan et al. 2017 or in Horowitz et al. 2019) are also expected to have different statistics and would also require custom calibration if they were to be used in a cosmological analysis.

As demonstrated in this work, simulations provide a useful means to quantify the statistics of CMB lensing cluster mass observables. Future work, however, will be needed in order to improve upon and extend our results on several fronts, for this CMB lensing observable and for others. First, the impact of baryonic effects on our results is difficult to quantify. Simulations that incorporate baryonic effects may be useful in this respect, although currently the number of massive galaxy clusters produced in such simulations is not large enough in order to compete statistically with the results presented in this chapter. Indeed, the number of galaxy clusters from state-of-the art simulations with baryons is only around 10–100 (Barnes et al., 2016; Planelles et al., 2017; Truong et al., 2017; Henden et al., 2019), significantly lower than our number of clusters, around 10^4. Furthermore, the possible dependence of our results on cosmological parameters has yet to be determined. The impact of deviations from log-normality on the mass calibration of a cluster sample and, in turn, on the cosmological constraints drawn from such sample also remains to be investigated (potentially along the lines of, e.g., Shaw et al. 2010). We have argued that this is a negligible effect for the *Planck* galaxy clusters, but it may not be negligible for future CMB experiments, which will deliver higher signal-to-noise CMB lensing cluster mass measurements. The correlations of CMB lensing mass observables with other cluster observables (e.g., the SZ and X-ray signals) also need to be quantified; simulations with baryons may also be useful in this respect. Future CMB experiments such as CMB-S4 (Abazajian et al.,

2016) will enable SZ counts analyses in which the SZ–mass scaling relations will be able to be calibrated completely with CMB lensing masses alone to sub-percent accuracy (Louis and Alonso, 2017). If their full statistical power is to be realised without biases, assessing the impact of these potential issues and accurately determining biases, scatter and intrinsic correlations will be an essential step for these studies.

Chapter 5

Conclusions and outlook

5.1 Thesis conclusions

In this thesis we have presented a cosmological analysis of the number counts of the galaxy clusters in the *Planck* MMF3 cosmology sample across tSZ signal-to-noise, CMB lensing signal-to-noise, and redshift. This sample consists of 439 *Planck* tSZ-detected galaxy clusters, selected on tSZ signal-to-noise. We have used the tSZ and redshift measurements as obtained by the *Planck* collaboration, and we have measured the CMB lensing signal-to-noise values of most of the clusters in the sample (433, those with measured redshift) from *Planck* temperature data, calibrating our CMB lensing cluster mass observable with mock observations from an N-body simulation. The cluster abundance across tSZ signal-to-noise and redshift provides most of the constraining power in our analysis, whereas the CMB lensing signal-to-noise measurements set the absolute cluster mass scale, which we have parametrised through an SZ mass bias parameter, $1-b_{\rm SZ}$. Assuming a flat ΛCDM cosmological model, and forward-modelling our cluster data self-consistently, we have found $\sigma_8(\Omega_{\rm m}/0.33)^{0.25} = 0.765 \pm 0.035$, $\Omega_{\rm m} = 0.33 \pm 0.02$, $\sigma_8 = 0.76 \pm 0.04$, and $1-b_{\rm SZ} = 0.71 \pm 0.10$. These constraints on ΛCDM parameters are in agreement with constraints from power spectrum measurements from *Planck* and from other low-redshift and high-redshift cosmological probes, and the constraint on the SZ mass bias is similarly in agreement with a number of other constraints on this parameter. These constraints also demonstrate the power of CMB lensing measurements to calibrate the absolute mass scale in cluster counts analyses, which will certainly be exploited further in the future.

Our analysis can be seen as a revisit of the tSZ counts analysis with CMB lensing mass calibration presented in Planck 2015 results XXIV (2016). That analysis also used the MMF3 cosmology sample, and the 433 clusters in it with measured redshift were also

used as the mass calibration sample. However, unlike in our analysis, the cluster mass calibration was attained through a prior on the SZ mass bias parameter, $1/(1-b_{\rm SZ}) = 0.99 \pm 0.19$, which was obtained from *Planck* CMB lensing mass measurements of the clusters in the calibration sample. With such a prior, and assuming a flat ΛCDM cosmology, this analysis found $\Omega_{\rm m} = 0.32 \pm 0.03$, $\sigma_8 = 0.71 \pm 0.03$. The fact that the constraint on σ_8 is in tension with the *Planck* CMB-derived constraint on this parameter, $\sigma_8 = 0.8111 \pm 0.0060$ (*Planck* 2018 TT,TE,EE+lowE+lensing results; Planck 2018 results VI 2018), led us to reanalyse this cluster sample with CMB lensing mass calibration. We have argued that an unaccounted-for selection effect and a slightly different CMB lensing cluster mass observable are responsible for the difference between our results and those in Planck 2015 results XXIV (2016).

Beyond our main results, the cosmological constraints presented in Chapter 3, other conclusions can be drawn from this thesis. Chapter 2 reveals how difficult cluster CMB lensing measurements are with current data (e.g., how small the signal-to-noise per cluster is, and how important foreground and mean-field removal are), yet it also demonstrates, as already shown in Planck 2015 results XXIV (2016), that such measurements are possible for current *Planck* data, with our measurements constituting a $4.5\,\sigma$ detection of cluster CMB lensing. This chapter also highlights the importance of the choice of a good mass observable. We indeed concluded that the CMB lensing signal-to-noise is better-behaved as an observable than the mass estimate itself, the former being essentially insensitive to the choice of matched filter input angular scale and to the assumed cosmology.

In Chapter 3 we presented the novel likelihood with which we performed our cluster counts analysis. This likelihood takes as data the total number of clusters in the sample, their sky locations, and the tSZ signal-to-noise, CMB lensing signal-to-noise, and redshift values of each cluster. In it, the Poissonian nature of cluster counts is encapsulated in an overall Poisson term for the total number of clusters. This differs from the usual Poisson counts-in-cell approach, having the advantage of avoiding binning in data space, which can be rather arbitrary and can entail a loss of signal-to-noise. A second term in the likelihood accounts for the clusters' sky locations, and a third term for their 'mass data points', their measured tSZ signal-to-noise, CMB lensing signal-to-noise, and redshift values. In this 'mass data likelihood', the cluster observables are modelled self-consistently through a hierarchical model that accounts for possible biases in each observable and for intrinsic and observational scatter around them. In this chapter we highlighted the importance of this consistent modelling of

cluster observables: it is an essential ingredient to obtain unbiased results given the presence of selection effects.

Finally, in Chapter 4 we calibrated the bias and the scatter of our preferred cluster CMB lensing mass observable, the CMB lensing signal-to-noise, with mock observations obtained from an N-body simulation. We also quantified deviations from the commonly-assumed log-normality in the scatter. This chapter demonstrates how crucial it is to understand the response of a given cluster mass observable to realistic realistic mass distributions of galaxy clusters and their environments, with simulations being a convenient way to do this. In particular, we argued that the bias in the observable must be accurately determined, as it is degenerate with the cluster mass, and therefore with parameters such as $\Omega_{\rm m}$ and σ_8. The intrinsic scatter has to be quantified as well if accurate cosmological constraints are to be obtained. In addition, in Chapter 4 we emphasised that these parameters must be determined for the precise specifications of the observable as it is used in the real analysis. As we illustrated quantitatively, changes in, for example, the convergence model used in the matched filter (e.g., NFW profile vs. NFW profile + two-halo term), or in the experimental specifications, significantly affect the observable bias and scatter. This will be especially important for future cluster surveys (e.g., from CMB-S4), in which, in order to achieve all their constraining potential, the cluster mass scale will have to be determined to percent-level accuracy, which means that sub-percent-level calibration of the mass observables used to set this mass scale will have to be attained.

5.2 The coming years

As noted in Chapter 1, cluster cosmology is in a very exciting time, with competitive constraints currently being derived, and with a potential statistical tension to be addressed. A lot of effort is currently being put into analysing data from past and present surveys (e.g., RASS, HSC-SSP, DES, KiDS, *Planck*, SPT, ACT), a task that will continue in the future. This often involves comparing data from observations at different frequencies in search of possible systematic biases. An example of this is the comparison of *Planck* SZ-selected clusters with X-ray selected clusters in Rossetti et al. (2016), Rossetti et al. (2017), and Andrade-Santos et al. (2017), which suggests that X-ray surveys preferentially detect relaxed, centrally-peaked clusters, and which, as a consequence, raises questions about how representative past X-ray samples are (Pratt et al., 2019). Comparing, and also combining, data from surveys at different frequencies will become increasingly important in the future. The synergy between observations

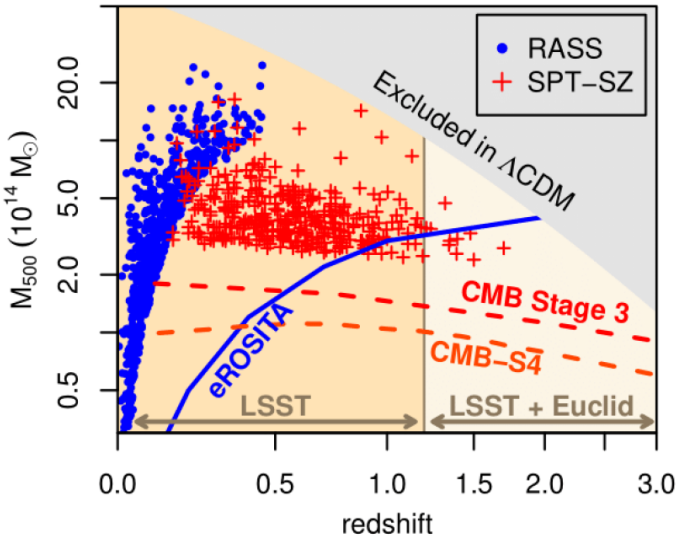


Fig. 5.1 Discovery space of some present and future galaxy cluster surveys, across cluster mass and redshift. Figure credit: Steven W. Allen slides from the KICC 10th Anniversary Symposium; `kicc.cam.ac.uk/files/Cambridge_KICC_Sep19_allen.pdf`.

at different frequencies should be exploited as much as possible in order to make the most of the large amount of high-quality data that upcoming experiments are going to produce over the next decade.

This future data will indeed come from different parts of the electromagnetic spectrum. The X-ray experiment eROSITA, launched in July 2019, will produce eight X-ray all-sky surveys with more than 20 times the sensitivity of RASS, enabling the detection of around 10^5 galaxy clusters out to $z \simeq 2$ (Pillepich et al., 2012), with precise temperature measurements for about 2000 of them (Borm et al., 2014). In parallel, LSST, a ground-based optical survey that will scan 20 000 deg^2 of the southern sky, observing for 10 years from its expected first light in 2020 (Ivezic et al., 2008), and *Euclid*, a space-based optical and near-infrared telescope that will scan 15 000 deg^2 over six years from its expected launch in 2022, will provide excellent optical observations. From them, around 10^5 galaxy clusters will be detected out to redshifts $z \simeq 1.2$ (LSST) and $z \simeq 2$ (*Euclid*; Sartoris et al. 2016), from which, among many other things, exquisite galaxy weak-lensing mass estimates will be obtained. In the mm, SPT-3G, the third-generation South Pole Telescope experiment, will detect around 5 000 clusters through their tSZ signature (Benson et al., 2014), and the Simons Observatory (SO), a CMB experiment based in the Atacama desert and with an expected first light in 2021, will find around 16 000 of them (Simons Observatory Collaboration, 2019). Towards the end of the decade, CMB-S4, a planned ground-based experiment with sites in

the Atacama desert and the South Pole, should increase this figure to around 10^5 (Abazajian et al., 2016).

These next-decade experiments will therefore detect most of the clusters in the observable Universe, which are expected to number around 10^5 (although cluster cosmology is, of course, far from being cosmic-variance-limited!). The regions of the cluster mass–redshift plane expected to be probed by these experiments are shown in Figure 5.1, along with typical RASS and SPT samples. The discovery space is huge, and, as expected, the different experiments are highly complementary.

Galaxy cluster analyses from these upcoming experiments will tighten the constraints on ΛCDM parameters such as Ω_m and σ_8, and will also probe the nature of dark energy, the masses of neutrinos, and primordial non-Gaussianity (see, e.g., Pillepich et al. 2012; Sartoris et al. 2016; Abazajian et al. 2016; Simons Observatory Collaboration 2019). Given their great constraining potential, keeping systematic effects under control will be more important than ever, and accurate cluster mass determination will therefore be crucial. In this respect, galaxy weak-lensing masses, from, e.g., LSST and *Euclid*, will probably continue to be the most robust anchors of the absolute cluster mass scale. However, significant progress is also expected in understanding deviations from hydrostatic equilibrium and the non-thermal pressure sources that bias hydrostatic mass estimates. A large contribution towards this effort will come from observations by XRISM (XRISM Science Team, 2020) and Athena. These two X-ray observatories, with expected launch dates in 2022 and 2031, respectively, will have very good spectral resolution, which will allow the direct measurement of turbulent and bulk motions in the inner regions of clusters (Ota et al., 2018; Roncarelli et al., 2018). Non-thermal pressure support will also be probed in outer regions by CMB experiments such as SO and CMB-S4. Finally, CMB lensing cluster mass determination will also benefit from this new generation of experiments.

Indeed, SO and CMB-S4 are expected to enable CMB lensing reconstruction with signal-to-noise ratio per lensing mode greater than unity up to $L \simeq 300$ and $L \simeq 1000$, respectively (Simons Observatory Collaboration, 2019; Abazajian et al., 2016). In SO, temperature and polarisation will roughly contribute in equal parts to the total lensing signal-to-noise ratio, although temperature will dominate the reconstruction on small scales important for cluster mass estimation, whereas polarisation will dominate in CMB-S4. For a sample of 1000 clusters, cluster masses will be obtained with few-percent precision (the specific expected precision depending on the method used), which means that sub-percent-level CMB lensing absolute mass calibration of future SZ cluster samples should be possible (Louis and Alonso, 2017), most effectively in combination

with weak-lensing masses. In addition, cluster CMB lensing is especially useful at high redshift, where it is virtually the only way to obtain robust mass estimates, and this region of observation space will be explored at an unprecedented level over the next decade (see Figure 5.1). Galaxy cluster CMB lensing, of which this thesis makes an early use, will therefore play an increasingly important role in cosmology in the years to come.

Appendix A

Effect of miscentering on cluster mass measurements

In this appendix we provide further details of the bias caused by miscentering in our lensing cluster mass estimates. We also propose a method to account for miscentering bias more carefully in a Bayesian analysis of cluster data.

A.1 Miscentring bias

As discussed in Section 2.4, we can build an estimator of a cluster's mass, $\hat{M}_{500}$, or, equivalently, of its signal-to-noise, $p_{\rm obs}$, by matched-filtering the cluster's reconstructed convergence profile with an appropriate template. This estimator is unbiased if the template matches exactly the true cluster's profile. This condition also requires that the template is placed exactly at the centre of the true profile. Here we consider the case in which the template is placed at a cluster's estimated centre, $\boldsymbol{x}_{\rm est}$, which is offset from the true centre, $\boldsymbol{x}_{\rm true}$, by a certain angle $\boldsymbol{d}$, i.e., $\boldsymbol{x}_{\rm true} = \boldsymbol{x}_{\rm est} + \boldsymbol{d}$. Assuming that otherwise the template is exactly the same as the true cluster's profile, the expected value of $p_{\rm obs}$ can be written as

$$\langle p_{\rm obs} \rangle (\boldsymbol{d}) = \mathscr{N} \int \frac{d^2\boldsymbol{L}}{2\pi} \frac{|\kappa(\boldsymbol{L})|^2}{N_\kappa(\boldsymbol{L})} e^{-i\boldsymbol{L}\cdot\boldsymbol{d}}, \tag{A.1}$$

where κ is the true profile and $\mathscr{N}$ is an appropriate normalisation. Expanding the phase term in a power series, only the even powers of $\boldsymbol{d}$ contribute since $\kappa(\boldsymbol{L}) = \kappa^*(-\boldsymbol{L})$ because the convergence is real. If we further assume circular symmetry for the convergence profile, so that $\kappa(\boldsymbol{L})$ only depends on $L = |\boldsymbol{L}|$, we can write the fractional

bias due to miscentering by $\boldsymbol{d}$ as

$$\frac{\langle p_{\rm obs}\rangle(\boldsymbol{d}) - \langle p_{\rm obs}\rangle(\boldsymbol{0})}{\langle p_{\rm obs}\rangle(\boldsymbol{0})} = \sum_{n\ \rm even} (-1)^{n/2} f^{(n)} d^n. \tag{A.2}$$

Here, the $f^{(n)}$ are coefficients quantifying the fractional bias at each order, which can be written as

$$f^{(n)} = \frac{1}{2^n[(n/2)!]^2} \left(\int dL\, L \frac{|\kappa(\boldsymbol{L})|^2}{N_\kappa(\boldsymbol{L})} \right)^{-1} \int dL\, L^{n+1} \frac{|\kappa(\boldsymbol{L})|^2}{N_\kappa(\boldsymbol{L})}. \tag{A.3}$$

We see that the leading-order bias is second order in the miscentering angle and is negative (while the next-to-leading-order term is positive). We note that the fractional bias in $p_{\rm obs}$ is also the fractional bias in $\hat{M}_{500}$, since the latter is related to $p_{\rm obs}$ through a factor unaffected by miscentering.

To estimate the size of the effect of miscentering, we can write the nth order contribution to the bias as

$$(-1)^{n/2} f^{(n)} d^n = (-1)^{n/2} \left(\frac{d}{\lambda^{(n)}} \right)^n, \tag{A.4}$$

where $\lambda^{(n)}$ is a typical angle given by $\lambda^{(n)} = (f^{(n)})^{-1/n}$. For a TT quadratic estimator reconstruction in an idealised *Planck*-like experiment with a Gaussian beam of FWHM = 5 arcmin and temperature noise levels of 45 μK-arcmin, we find $\lambda^{(2)} \approx 6$ arcmin and $\lambda^{(4)} \approx 7$ arcmin. For a cluster with $M_{500} = 0.5 \times 10^{15} M_\odot$ at $z = 0.3$ and a typical miscentering error of around 1 arcmin, the second-order fractional bias is around 10^{-2} (i.e., at the few percent level), while the fourth-order bias is around 10^{-4}.

A.2 Accounting for miscentering in a likelihood analysis

As explained in Section 3.4, in our likelihood analysis we model the miscentering bias with an effective lensing mass bias parameter, which we can marginalise over. Here we describe a more rigorous way to deal with miscentering bias in a likelihood analysis, although we leave implementation of this approach to future work.

If we knew the centring offset $\boldsymbol{d}$, for a cluster with true lensing signal-to-noise p_t the average of the observed signal-to-noise over other sources of observational scatter is

$$\langle p_\mathrm{obs}\rangle(\boldsymbol{d}) \approx p_\mathrm{t}\left(1+\sum_{n\ \mathrm{even}}(-1)^{n/2}f^{(n)}d^n\right). \tag{A.5}$$

Here, the bias coefficients $f^{(n)}$ should be evaluated at the appropriate cluster parameters (M_{500}, z, $\hat{\boldsymbol{n}}$, $1-b_\mathrm{CMBlens}$) in the hierarchical model. However, they are independent of $\boldsymbol{d}$, which proves useful for efficient marginalisation over $\boldsymbol{d}$.

Obviously, $\boldsymbol{d}$ is not known precisely, but experiments such as *Planck* do provide an estimate of the uncertainty in each reported cluster position. For a given cluster, let us assume that its position on the sky is described by a certain two-dimensional probability distribution, $P(\boldsymbol{d})$. This could be a two-dimensional Gaussian, for example. What we are interested in is the probability distribution followed by p_obs, $P(p_\mathrm{obs}|p_\mathrm{t}, M_{500}, z, \hat{\boldsymbol{n}})$, after the uncertainty due to miscentering has been taken into account. This is given by

$$P(p_\mathrm{obs}|p_\mathrm{t}, M_{500}, z, \hat{\boldsymbol{n}}) = \int d^2\boldsymbol{d}\, P(p_\mathrm{obs}|p_\mathrm{t}, M_{500}, z, \hat{\boldsymbol{n}}, \boldsymbol{d})P(\boldsymbol{d}), \tag{A.6}$$

where $P(p_\mathrm{obs}|p_\mathrm{t}, M_{500}, z, \hat{\boldsymbol{n}}, \boldsymbol{d})$ is the conditional distribution for p_obs given the centring error $\boldsymbol{d}$. Retaining the Gaussian approximation for the other observational sources of scatter, this conditional distribution is simply a Gaussian with mean given by Eq. (A.5) and unit variance.

The series expansion in Eq. (A.5) allows $P(p_\mathrm{obs}|p_\mathrm{t}, M_{500}, z, \hat{\boldsymbol{n}})$ to be evaluated efficiently. We note that if indeed we take $P(p_\mathrm{obs}|p_\mathrm{t}, M_{500}, z, \hat{\boldsymbol{n}}, \boldsymbol{d})$ and $P(\boldsymbol{d})$ both to be Gaussians, and if we correct for miscentering at least to the lowest non-vanishing order (quadratic), then $P(p_\mathrm{obs}|p_\mathrm{t}, M_{500}, z, \hat{\boldsymbol{n}})$ will no longer be Gaussian and will depend (weakly) on M_{500}, z, and $\hat{\boldsymbol{n}}$ through the $f^{(n)}$. Thus, the method outlined here allows to account for miscentering uncertainty in a rigorous way at the expense of complicating the statistics of our mass observable. We note that a similar miscentering argument should apply to the SZ signal-to-noise measurements, which should have common miscentering errors with the CMB lensing signal-to-noise measurements. Accounting properly for miscentering would therefore imply marginalising over $\boldsymbol{d}$ simultaneously in both observables.

Appendix B

On the selection bias

As mentioned in Section 3.3, the scaling-relation-derived masses of the galaxy clusters in the MMF3 cosmology sample are biased due to the fact that the sample is obtained through a cut-off on the SZ signal-to-noise. The CMB lensing masses are also biased, since they correlate with the scaling-relation-derived masses. This kind of selection effect has already been discussed in the literature; see, e.g., Allen et al. (2011); Sereno and Ettori (2015); Nagarajan et al. (2019). In this appendix, we give a qualitative explanation of the origin of this bias, illustrating it with a toy model. We also give a mathematical proof of its existence under more general conditions.

B.1 Qualitative explanation

The MMF3 cosmology sample, consisting of 439 galaxy clusters, was selected from the whole sample of clusters detected by *Planck* through a cut-off in the observed SZ signal-to-noise, $q_{\rm obs} > 6$. As explained in the main text, $q_{\rm obs}$ is a noisy realisation of the mean signal-to-noise at given true mass, M_{500}, and redshift, z, $\bar{q}(M_{500}, z)$. Thus, by imposing a cut-off in $q_{\rm obs}$ in order to select our sample, at a given $\bar{q}(M_{500}, z)$, clusters with a large $q_{\rm obs}$ are being preferentially selected. Even if before imposing the selection criterion $q_{\rm obs}$ was an unbiased estimator of $\bar{q}(M_{500}, z)$, after selection $q_{\rm obs}$ becomes a biased estimator of $\bar{q}(M_{500}, z)$. One case use $q_{\rm obs}$ to obtain a cluster mass estimate, M_{Y_z}, through the SZ–mass scaling relations. Since $q_{\rm obs}$ is preferentially large at given M_{500} and z, so will be M_{Y_z}. It follows that close to the selection threshold, M_{Y_z} is a biased estimator of the mean SZ mass at given M_{500}, which is the mass that enters the scaling relations.

This effect can be easily illustrated with simulated data generated from a toy model, such as the data points plotted in Fig. B.1. Similar plots can be found in, e.g., Allen

et al. (2011) and Nagarajan et al. (2019). In order to produce these simulated data points, we sample from the halo mass function times the volume element (Eq. 3.15) to obtain values of M_{500} and, then, for each cluster, we generate a value of its 'SZ mass', M_{SZ}, as a realisation of a log-normal process centred on $(1 - b_{\rm SZ})M_{500}$, where $1 - b_{\rm SZ} = 0.7$. (The blue line in Fig. B.1 shows the relation $M_{\rm SZ} = (1 - b_{\rm SZ})M_{500}$.) We then impose a cut-off on $M_{\rm SZ}$ (green line) in order to select our sample (blue points), discarding all the clusters whose value of $M_{\rm SZ}$ falls below a certain mass threshold (red points). We remind the reader that in our real cluster sample, the cut-off is in $q_{\rm obs}$ rather than in $M_{\rm SZ}$, but a simpler cut-off in $M_{\rm SZ}$ is sufficient to understand the origin of the selection effect. It can clearly be observed that, at a given M_{500}, clusters with a large $M_{\rm SZ}$ are being selected. It can also be observed that if $1 - b_{\rm SZ}$ were to be estimated by dividing $M_{\rm SZ}$ by M_{500} for each cluster and averaging over the selected sample, the estimate would be biased high. A similar bias would arise if a straight line of unit gradient were fit to the selected clusters in log–log space, with the intercept $\ln(1 - b_{\rm SZ})$ being biased high.

Let us now consider a second estimate of M_{500}, such as a lensing mass, and assume that, before any selection criteria are imposed, it is an unbiased estimate of M_{500}. If the scatter in the lensing masses is uncorrelated with that in the SZ mass estimates (M_{Y_z}, or $M_{\rm SZ}$ in our toy model), the lensing mass remains unbiased after the cluster selection, since it does not intervene in the selection process. However, if there is correlated scatter, the lensing masses will also be biased to an extent that depends on the amount of correlation and on their scatter. As noted in the main text, lensing masses are expected to correlate positively with SZ masses due to the fact that they are both derived from quantities integrated along the line of sight; indeed, this has been observed in simulations (Shirasaki et al., 2016). We therefore expect the lensing masses also to be biased to some extent due to this selection effect.

B.2 An analytical proof of the existence of the bias

We now give a general mathematical proof of the existence of a bias in an observable arising from the selection of a sample through a cut-off in that observable.

Let M be a random variable with probability distribution function $P(M)$. We assume that $P(M)$ only takes non-zero values in a region where $M > 0$. We identify M with the true value of a physical quantity that we want to measure, e.g., the mass of a galaxy cluster. Let us now consider some scatter on that quantity, due to, e.g., statistical or instrumental uncertainty. Let us denote by D the variable corresponding

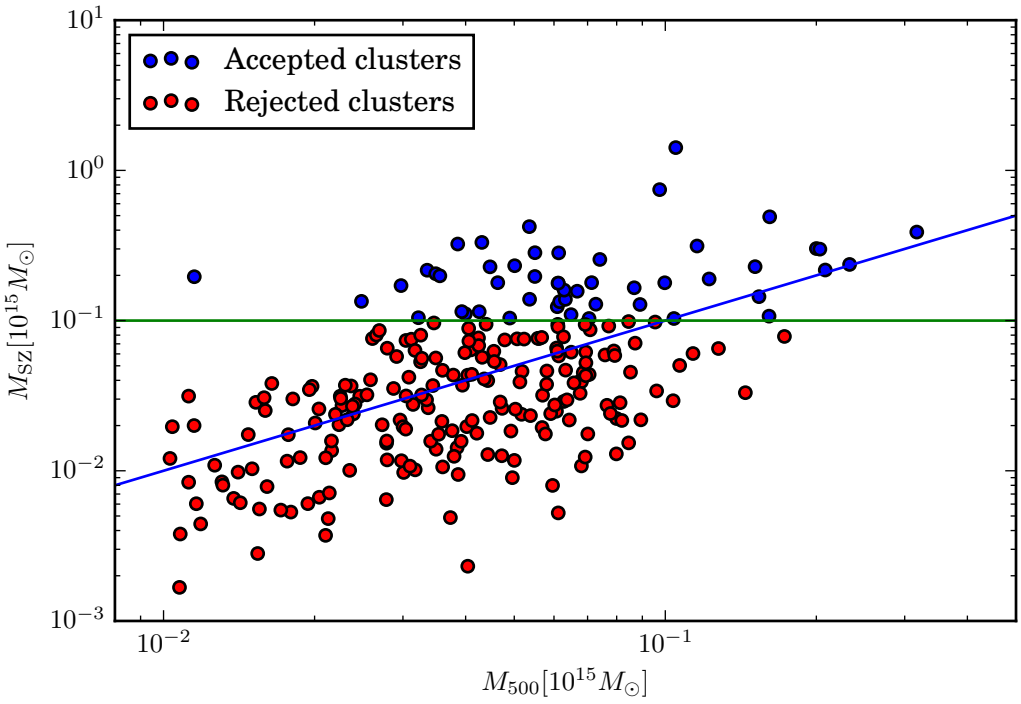


Fig. B.1 Simulated clusters for our toy model. The accepted clusters, with SZ masses exceeding the selection threshold (green line) are shown in blue, while the rejected clusters are shown in red. The blue line is the (population) mean $M_{\rm SZ}$ at a given M_{500}, i.e., $(1 - b_{\rm SZ})M_{500}$.

to the measured values of M, which, for a given M, has probability density $P(D|M)$. We assume that D is an unbiased measurement of M, i.e.,

$$\int_{-\infty}^{\infty} dD\, DP(D|M) = M. \tag{B.1}$$

The probability density of D follows $P(D)$, where

$$P(D) = \int_{-\infty}^{\infty} dM\, P(D|M)P(M). \tag{B.2}$$

Suppose that we make a series of measurements of the physical quantity D, i.e., that we sample from $P(D)$, and that we impose a selection threshold retaining only those elements with $D > D_t$. The values of D selected this way have a probability density $P(D|\text{in})$, where "in" denotes the inclusion condition, with

$$P(D|\text{in}) = \frac{P(\text{in}|D)P(D)}{P(\text{in})}. \tag{B.3}$$

Here, $P(\text{in}|D)$ is the probability of inclusion in our sample given a certain D, and is simply a unit step function centred at $D = D_t$, and $P(\text{in})$ is a normalisation factor quantifying the total probability of inclusion. Let us denote the expected value of D given that it is included in the sample by μ_D, so that

$$\mu_D = \int_{-\infty}^{\infty} dD\, DP(D|\text{in}). \tag{B.4}$$

We can also consider the distribution of M given inclusion, $P(M|\text{in})$. Let us denote the associated expectation value of M by μ_M,

$$\mu_M = \int_{-\infty}^{\infty} dM\, MP(M|\text{in}). \tag{B.5}$$

We now claim that $\mu_D > \mu_M$, i.e., the expected value of the observations of the physical quantity of interest (D) is greater than the mean of their true values (M) for a sample selected through a cutoff on D.

The proof is straightforward. Let us write μ_D as

$$\begin{aligned}\mu_D &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} dDdM\, DP(D|M,\text{in})P(M|\text{in}) \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} dDdM\, D\frac{P(\text{in}|D,M)P(D|M)}{P(\text{in}|M)}P(M|\text{in}).\end{aligned} \tag{B.6}$$

We have that $P(\text{in}|D, M) = P(\text{in}|D)$, and we can write $P(\text{in}|M)$ as

$$P(\text{in}|M) = \int_{-\infty}^{\infty} dD\, P(\text{in}|D, M)P(D|M) = \int_{D_t}^{\infty} dD\, P(D|M). \tag{B.7}$$

It follows that we can write μ_D as

$$\mu_D = \int_{-\infty}^{\infty} dM\, F(M)P(M|\text{in}), \tag{B.8}$$

where

$$F(M) = \frac{\int_{D_t}^{\infty} dD\, DP(D|M)}{\int_{D_t}^{\infty} dD\, P(D|M)}. \tag{B.9}$$

Now, $F(M)$ can be understood as the expected value of D under a probability distribution that is proportional to $P(D|M)$ for $D > D_t$, and zero otherwise (with the term in the denominator on the right-hand side providing the appropriate normalisation). Since the mean of D under the non-truncated $P(D|M)$ is M (Eq. B.1), and the truncation at D_t is a lower truncation, it follows that $F(M) > M$ and so, from Eq. (B.8), that $\mu_D > \mu_M$. We can write the bias as

$$\mu_D - \mu_M = \int_{-\infty}^{\infty} dM\, [F(M) - M]\, P(M|\text{in}). \tag{B.10}$$

To summarise, if a sample is selected through a cut-off on the observed, noisy values of an observable, D, the expected value of the samples of D is larger than the expectation of the true values of the observable for the sample, M. It can easily be seen that the bias disappears (for $M > D_t$) in the limit of no scatter, for which $P(D|M) = \delta(D - M)$.

A graphical illustration of the different distributions involved in this discussion can be seen in Fig. B.2, where $P(M)$, $P(D)$, $P(M|\text{in})$, and $P(D|\text{in})$ are plotted for a toy model in which $P(M)$ and $P(D|M)$ are both Gaussians, and with the cut-off at $D_t = 0.8$. It should be apparent graphically that the expected value of D in the distribution $P(D|\text{in})$ is larger than that of M in $P(M|\text{in})$.

Let us now consider a second proxy of the observable M, which we denote by D'. The joint probability density of D and D' at a given M is $P(D, D'|M)$ and we assume that the expected values of both D and D' are M: both measurements are unbiased, but they may be correlated. Let us also denote with $P(D, D'|\text{in})$ the joint probability of D and D' given our inclusion condition, which we remind is set by D; D' can be understood as a follow-up measurement of M for the elements of our D-selected sample.

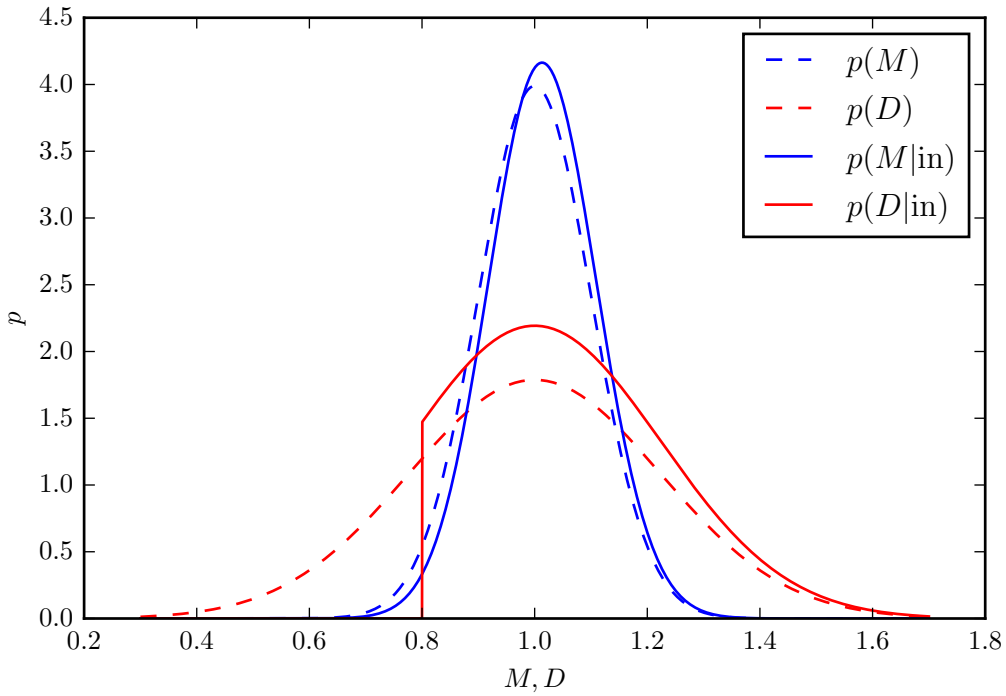


Fig. B.2 Probability density functions $P(M)$, $P(D)$, $P(M|\text{in})$, and $P(D|\text{in})$ for a toy model in which $P(M)$ and $P(D|M)$ are Gaussians, and with the cut-off at $D_t = 0.8$.

In this case, D is still a biased proxy of M; the situation is exactly the same as given above as is easily seen by marginalising over D'. On the other hand, the expected value of D' in the sample, $\mu_{D'}$, can be written as

$$\begin{aligned}\mu_{D'} &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} dDdD'\, D'P(D,D'|\mathrm{in}) \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} dDdD'dM\, D'P(D,D'|M,\mathrm{in})P(M|\mathrm{in}) \qquad \text{(B.11)} \\ &= \int_{-\infty}^{\infty} dM\, G(M)P(M|\mathrm{in}), \qquad \text{(B.12)}\end{aligned}$$

where

$$G(M) = \frac{\int_{-\infty}^{\infty} dD' \int_{D_t}^{\infty} dD\, D'P(D,D'|M)}{\int_{-\infty}^{\infty} dD' \int_{D_t}^{\infty} dD\, P(D,D'|M)}. \qquad \text{(B.13)}$$

If D and D' are statistically independent, $P(D,D'|M) = P(D|M)P(D'|M)$, so $G(M) = M$, and, therefore, $\mu_{D'} = \mu_M$. This can be seen graphically in Fig. B.3, which illustrates a contour of equal probability density of $P(D,D'|M)$ in the D–D' plane in this case of statistical independence. The quantity $G(M)$ can be understood as the expected value of D' for a probability distribution function proportional to $P(D,D'|M)$ for $D > D_t$ and zero otherwise. As can be seen in the figure, for this truncated distribution the mean of D is greater than M [hence $F(M) > M$], but the mean of D' is unchanged from M [so $G(M) = M$].

On the other hand, if D and D' are not statistically independent, in general $\mu_{D'} \neq \mu_M$. This is illustrated in the lower panel of Fig. B.3, where a positive correlation between D and D' is assumed. It can be observed that, now, for the truncated distribution both the means of D and D' exceed M, so that $\mu_D > \mu_M$ and $\mu_{D'} > \mu_M$. Analogously, a negative correlation leads to $\mu_D > \mu_M$ and $\mu_{D'} < \mu_M$.

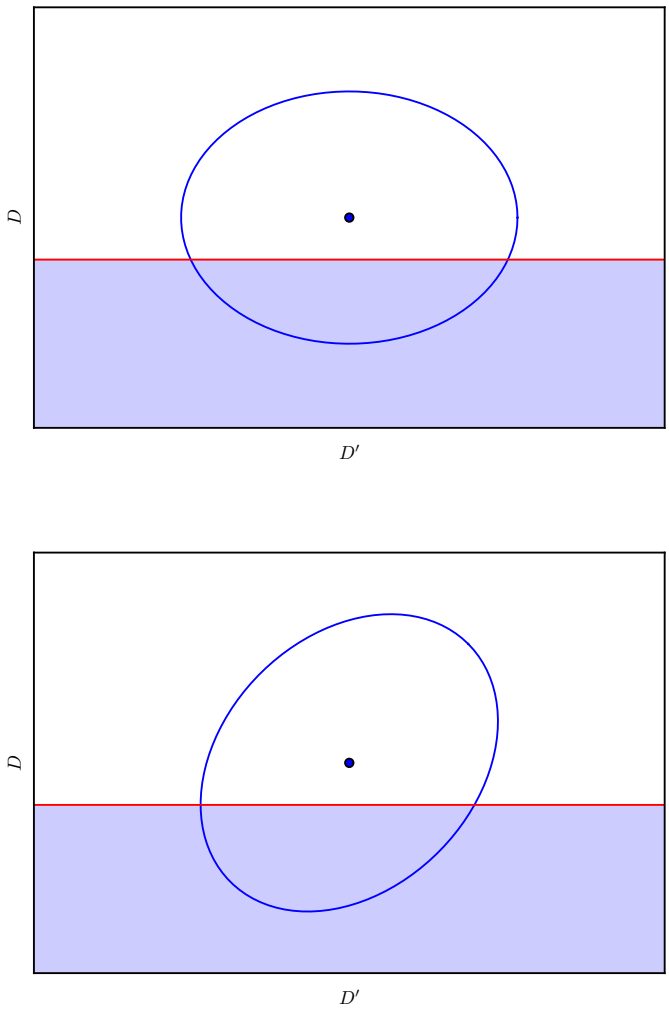


Fig. B.3 Illustrations of a contour of constant probability density for the distribution $P(D, D'|M)$ in the case of no correlation (top), and positive correlation (bottom), between D and D'. In both cases, the red line depicts the selection cut-off, with the shaded region being the region that is truncated. The contours are centred on the point (M, M), marked with a dot.

Appendix C

Relating our likelihood to a Poisson counts likelihood

In this appendix we demonstrate how the likelihood used in this paper, which is described in detail in Section 3.4, is mathematically equivalent to a Poisson counts likelihood in the $q_{\rm obs}$–$p_{\rm obs}$–z–$\hat{\boldsymbol{n}}$ space for cells of sufficiently small size. We also show how it can be reduced to a likelihood similar to the one used in the *Planck* SZ counts analysis (Planck 2015 results XXIV, 2016) by suitable marginalisation.

Let us consider an M-dimensional data space for each galaxy cluster; in our case, $M = 5$. We continue to assume that clusters can be treated independently of each other (i.e., effects of spatial clustering are negligible), so that our cluster sample can be considered a realisation of a Poisson process in this data space. One way to describe the sample statistically is via a Poisson counts approach.

In this approach, we first divide the M-dimensional (continuous) data space into a set of $N_{\rm bin}$ cells, covering the whole region of data space where there is a non-zero probability for a cluster to be located. We assume that each cell has equal volume Ω. Since the clusters are assumed to be statistically independent from each other, the number of clusters with data falling within cell I, n_I, follows a Poisson distribution, and the joint probability distribution for the sample, now described by the cell occupation numbers $\boldsymbol{n} = (n_1, \ldots, n_{N_{\rm bin}})$ is simply given by the product of the Poisson distributions of each cell,

$$P(\boldsymbol{n}) = \prod_I \frac{\bar{n}_I^{n_I} e^{-\bar{n}_I}}{n_I!}. \tag{C.1}$$

Here, $\bar{n}_I$ is the expected number of clusters in cell I. We can write $\bar{n}_I = P_I \bar{N}$, where P_I is the probability for a single cluster to lie in cell I given that it is passes the selection criterion to be in the sample, and $\bar{N} = \sum_I \bar{n}_I$ is the expected total number of clusters

in the sample. We can then rewrite the likelihood as

$$P(\boldsymbol{n}) = \frac{\bar{N}^N e^{-\bar{N}}}{N!} N! \prod_I \frac{P_I^{n_I}}{n_I!}, \tag{C.2}$$

where $N = \sum_I n_I$ is the total number of clusters in the sample. The first term in Eq. (C.2) is a Poisson distribution for N, which we shall refer to as $P(N)$. The second term is a multinomial distribution for $\boldsymbol{n}$ with N trials, that is, the probability distribution followed by $\boldsymbol{n}$ given a fixed N. We shall refer to it as $P(\boldsymbol{n}|N)$. We have therefore factored a product of Poisson distributions across the cells into a Poisson distribution for the total number of clusters and a distribution for the particular configuration of cluster data points at a fixed total number of clusters,

$$P(\boldsymbol{n}) = P(\boldsymbol{n}|N)P(N). \tag{C.3}$$

Let us now consider how this Poisson counts likelihood relates to the unbinned likelihood used in the rest of this paper. Let us denote points in the M-dimensional data space by $\boldsymbol{x}$ and the data corresponding to the ith cluster as $\boldsymbol{x}_i$. Also, let $P(\boldsymbol{x}_i|\mathrm{in})$ be the probability density for a cluster to have data $\boldsymbol{x}_i$ given that it passes the sample-selection criterion. A given realisation of the cluster sample consists of specifying N and the set $\underline{D}' = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N)$. As the cluster labels are arbitrary, there are $N!$ permutations that give equivalent samples of clusters. Each permutation gives the same cell occupation numbers $\boldsymbol{n}$, which are got from counting the number of clusters n_I that have their data $\boldsymbol{x}_i$ within cell I. Now consider the limit in which the cells that are small enough that the variation of $P(\boldsymbol{x}|\mathrm{in})$ across the cell can be ignored. If some central value of $\boldsymbol{x}$ in the Ith cell is denoted by $\boldsymbol{x}_I$, for a given sample of N clusters we have

$$\prod_{I=1}^{N_{\mathrm{bin}}} P_I^{n_I} \approx \prod_{I=1}^{N_{\mathrm{bin}}} [P(\boldsymbol{x}_I|\mathrm{in})\Omega]^{n_I} \approx \prod_{i=1}^{N} [P(\boldsymbol{x}_i|\mathrm{in})\Omega] = P(\underline{D}'|N)\Omega^N. \tag{C.4}$$

Thus, in this limit, we can rewrite the cells likelihood, Eq. (C.2), as

$$P(\boldsymbol{n}) \propto P(\underline{D}'|N)P(N), \tag{C.5}$$

where the proportionality factor, which accounts for the change in measure between the two data descriptions, is cosmology independent.

Equation (C.5) has the same form as our likelihood, presented in Section 3.4. There, we separated the angular positions of the N clusters, $\hat{\underline{\boldsymbol{n}}}$, from the other cluster

observables, $\underline{D}$, so that $\underline{D}' = (\underline{D}, \underline{\hat{\boldsymbol{n}}})$. Indeed, Eq. (3.3) can be written as

$$P\left(N, \underline{\hat{\boldsymbol{n}}}, \underline{D}\right) = P(\underline{D}, \underline{\hat{\boldsymbol{n}}}|N)P(N) = P(\underline{D}'|N)P(N), \tag{C.6}$$

where $P(\underline{D}, \underline{\hat{\boldsymbol{n}}}|N)$ is just the product of the mass data likelihood, $\mathscr{L}_1$ (see Section 3.4.1), and the sky location likelihood, $\mathscr{L}_2$ (see Section 3.4.2). Our likelihood is therefore equivalent to a Poisson counts likelihood in the D–$\hat{\boldsymbol{n}}$ space (or, more explicitly, the q_{obs}–p_{obs}–z–$\hat{\boldsymbol{n}}$ space), with cells of sufficiently small size. This also implies that if we were to marginalise our likelihood over all the p_{obs} and $\hat{\boldsymbol{n}}$ variables, we would obtain a likelihood equivalent to a Poisson counts likelihood in the q_{obs}–z plane (with cells of sufficiently small size). This is similar to the *Planck* SZ counts likelihood used in Planck 2015 results XXIV (2016), except there cells with a fixed finite size are used.

References

Abazajian, K. N., Adshead, P., Ahmed, Z., Allen, S. W., Alonso, D., Arnold, K. S., Baccigalupi, C., Bartlett, J. G., Battaglia, N., Benson, B. A., et al. (2016). CMB-S4 Science Book, First Edition. *ArXiv e-prints*.

Abbott, T. M. C., Alarcon, A., Allam, S., Andersen, P., Andrade-Oliveira, F., Annis, J., Asorey, J., Avila, S., Bacon, D., Banik, N., et al. (2019). Cosmological constraints from multiple probes in the dark energy survey. *Phys. Rev. Lett.*, 122:171301.

Abdullah, M. H., Klypin, A., and Wilson, G. (2020). Cosmological Constraint on Ω_m and σ_8 from Cluster Abundances using the `GalWCat19` Optical-Spectroscopic SDSS Catalog. *arXiv e-prints*, page arXiv:2002.11907.

Abell, G. O. (1958). The Distribution of Rich Clusters of Galaxies. ApJS, 3:211.

Allen, S. W., Evrard, A. E., and Mantz, A. B. (2011). Cosmological Parameters from Observations of Galaxy Clusters. ARA&A, 49:409–470.

Allen, S. W., Rapetti, D. A., Schmidt, R. W., Ebeling, H., Morris, R. G., and Fabian, A. C. (2008). Improved constraints on dark energy from Chandra X-ray observations of the largest relaxed galaxy clusters. MNRAS, 383(3):879–896.

Allen, S. W., Schmidt, R. W., Ebeling, H., Fabian, A. C., and van Speybroeck, L. (2004). Constraints on dark energy from Chandra observations of the largest relaxed galaxy clusters. MNRAS, 353(2):457–467.

Allen, S. W., Schmidt, R. W., Fabian, A. C., and Ebeling, H. (2003). Cosmological constraints from the local X-ray luminosity function of the most X-ray-luminous galaxy clusters. MNRAS, 342(1):287–298.

Alpher, R. A., Bethe, H., and Gamow, G. (1948). The Origin of Chemical Elements. *Physical Review*, 73(7):803–804.

Alpher, R. A., Follin, J. W., and Herman, R. C. (1953). Physical Conditions in the Initial Stages of the Expanding Universe. *Physical Review*, 92(6):1347–1361.

Alpher, R. A. and Herman, R. (1948). Evolution of the Universe. Nature, 162(4124):774–775.

Ameglio, S., Borgani, S., Pierpaoli, E., Dolag, K., Ettori, S., and Morandi, A. (2009). Reconstructing mass profiles of simulated galaxy clusters by combining Sunyaev-Zeldovich and X-ray images. MNRAS, 394(1):479–490.

Anderes, E., Wandelt, B. D., and Lavaux, G. (2015). Bayesian Inference of CMB Gravitational Lensing. ApJ, 808:152.

Andrade-Santos, F., Jones, C., Forman, W. R., Lovisari, L., Vikhlinin, A., van Weeren, R. J., Murray, S. S., Arnaud, M., Pratt, G. W., Démoclès, J., Kraft, R., Mazzotta, P., Böhringer, H., Chon, G., Giacintucci, S., Clarke, T. E., Borgani, S., David, L., Douspis, M., Pointecouteau, E., Dahle, H., Brown, S., Aghanim, N., and Rasia, E. (2017). The Fraction of Cool-core Clusters in X-Ray versus SZ Samples Using Chandra Observations. ApJ, 843(1):76.

Angulo, R. E., Springel, V., White, S. D. M., Jenkins, A., Baugh, C. M., and Frenk, C. S. (2012). Scaling relations for galaxy clusters in the Millennium-XXL simulation. MNRAS, 426:2046–2062.

Arnaud, M., Pointecouteau, E., and Pratt, G. W. (2007). Calibration of the galaxy cluster M{500}-Y{X} relation with XMM-Newton. A&A, 474(3):L37–L40.

Arnaud, M., Pratt, G. W., Piffaretti, R., Böhringer, H., Croston, J. H., and Pointecouteau, E. (2010). The universal galaxy cluster pressure profile from a representative sample of nearby systems (REXCESS) and the Y_{SZ} - M_{500} relation. A&A, 517:A92.

Bahcall, N. A. and Soneira, R. M. (1983). The spatial correlation function of rich clusters of galaxies. ApJ, 270:20–38.

Bahé, Y. M., McCarthy, I. G., and King, L. J. (2012). Mock weak lensing analysis of simulated galaxy clusters: bias and scatter in mass and concentration. MNRAS, 421:1073–1088.

Baltz, E. A., Marshall, P., and Oguri, M. (2009). Analytic models of plausible gravitational lens potentials. J. Cosmology Astropart. Phys., 2009(1):015.

Bardeen, J. M. (1980). Gauge-invariant cosmological perturbations. Phys. Rev. D, 22(8):1882–1905.

Bardeen, J. M., Bond, J. R., Kaiser, N., and Szalay, A. S. (1986). The Statistics of Peaks of Gaussian Random Fields. ApJ, 304:15.

Barnes, D. J., Kay, S. T., Henson, M. A., McCarthy, I. G., Schaye, J., and Jenkins, A. (2016). The redshift evolution of massive galaxy clusters in the MACSIS simulations. *Monthly Notices of the Royal Astronomical Society*, 465(1):213–233.

Bartelmann, M. (2010). TOPICAL REVIEW Gravitational lensing. *Classical and Quantum Gravity*, 27(23):233001.

Bartolo, N., Komatsu, E., Matarrese, S., and Riotto, A. (2004). Non-Gaussianity from inflation: theory and observations. Phys. Rep., 402(3-4):103–266.

Battaglia, N., Bond, J. R., Pfrommer, C., and Sievers, J. L. (2012). On the Cluster Physics of Sunyaev-Zel'dovich and X-Ray Surveys. I. The Influence of Feedback, Non-thermal Pressure, and Cluster Shapes on Y-M Scaling Relations. ApJ, 758(2):74.

Baumann, D. (2009). TASI Lectures on Inflation. *arXiv e-prints*, page arXiv:0907.5424.

Baxter, E. J., Keisler, R., Dodelson, S., et al. (2015). A Measurement of Gravitational Lensing of the Cosmic Microwave Background by Galaxy Clusters Using Data from the South Pole Telescope. ApJ, 806:247.

Baxter, E. J., Raghunathan, S., Crawford, T. M., Fosalba, P., Hou, Z., Holder, G. P., Omori, Y., Patil, S., Rozo, E., et al. (2017). A Measurement of CMB Cluster Lensing with SPT and DES Year 1 Data. J. Cosmology Astropart. Phys..

Baxter, E. J., Sherwin, B. D., and Raghunathan, S. (2019). Constraining the rotational kinematic Sunyaev-Zel'dovich effect in massive galaxy clusters. J. Cosmology Astropart. Phys., 2019(6):001.

Becker, M. R. and Kravtsov, A. V. (2011). On the Accuracy of Weak-lensing Cluster Mass Reconstructions. ApJ, 740:25.

Bennett, C. L., Banday, A. J., Gorski, K. M., Hinshaw, G., Jackson, P., Keegstra, P., Kogut, A., Smoot, G. F., Wilkinson, D. T., and Wright, E. L. (1996). Four-Year COBE DMR Cosmic Microwave Background Observations: Maps and Basic Results. ApJ, 464:L1.

Benoit-Lévy, A., Déchelette, T., Benabed, K., Cardoso, J.-F., Hanson, D., and Prunet, S. (2013). Full-sky CMB lensing reconstruction in presence of sky-cuts. A&A, 555:A37.

Benson, B. A., de Haan, T., Dudley, J. P., Reichardt, C. L., et al. (2013). Cosmological Constraints from Sunyaev-Zel'dovich-selected Clusters with X-Ray Observations in the First 178 deg^2 of the South Pole Telescope Survey. ApJ, 763(2):147.

Benson, B. A. et al. (2014). *SPT-3G: a next-generation cosmic microwave background polarization experiment on the South Pole telescope*, volume 9153 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, page 91531P.

Bianchini, F., Bielewicz, P., Lapi, A., Gonzalez-Nuevo, J., Baccigalupi, C., de Zotti, G., Danese, L., Bourne, N., Cooray, A., Dunne, L., and et al. (2015). Cross-correlation between the cmb lensing potential measured by planck and high-z submillimiter galaxies detected by the herschel-atlas survey. *The Astrophysical Journal*, 802(1):64.

BICEP2 Collaboration and Keck Array Collaboration (2018). Constraints on Primordial Gravitational Waves Using Planck, WMAP, and New BICEP2/Keck Observations through the 2015 Season. Phys. Rev. Lett., 121(22):221301.

Biffi, V., Borgani, S., Murante, G., Rasia, E., Planelles, S., Granato, G. L., Ragone-Figueroa, C., Beck, A. M., Gaspari, M., and Dolag, K. (2016). On the Nature of Hydrostatic Equilibrium in Galaxy Clusters. ApJ, 827(2):112.

Birkinshaw, M., Gull, S. F., and Hardebeck, H. (1984). The Sunyaev-Zeldovich effect towards three clusters of galaxies. Nature, 309:34–35.

Biviano, A. (2000). From Messier to Abell: 200 Years of Science with Galaxy Clusters. In *Constructing the Universe with Clusters of Galaxies*.

References

Biviano, A. and Girardi, M. (2003). The Mass Profile of Galaxy Clusters out to $\sim 2r_{200}$. ApJ, 585(1):205–214.

Blanchard, A. and Schneider, J. (1987). Gravitational lensing effect on the fluctuations of the cosmic background radiation. A&A, 184(1-2):1–6.

Bleem, L. E., Stalder, B., de Haan, T., et al. (2015). Galaxy Clusters Discovered via the Sunyaev-Zel'dovich Effect in the 2500-Square-Degree SPT-SZ Survey. ApJS, 216:27.

Bocquet, S., Dietrich, J. P., Schrabback, T., Bleem, L. E., Klein, M., et al. (2019). Cluster Cosmology Constraints from the 2500 deg^2 SPT-SZ Survey: Inclusion of Weak Gravitational Lensing Data from Magellan and the Hubble Space Telescope. ApJ, 878:55.

Bocquet, S., Heitmann, K., Habib, S., Lawrence, E., Uram, T., Frontiere, N., Pope, A., and Finkel, H. (2020). The Mira-Titan Universe. III. Emulation of the Halo Mass Function. *arXiv e-prints*, page arXiv:2003.12116.

Bocquet, S., Saro, A., Dolag, K., and Mohr, J. J. (2016). Halo mass function: baryon impact, fitting formulae, and implications for cluster cosmology. MNRAS, 456(3):2361–2373.

Böhringer, H., Schuecker, P., Guzzo, L., Collins, C. A., Voges, W., Cruddace, R. G., Ortiz-Gil, A., Chincarini, G., De Grandi, S., Edge, A. C., MacGillivray, H. T., Neumann, D. M., Schindler, S., and Shaver, P. (2004). The ROSAT-ESO Flux Limited X-ray (REFLEX) Galaxy cluster survey. V. The cluster catalogue. A&A, 425:367–383.

Böhringer, H., Voges, W., Huchra, J. P., McLean, B., Giacconi, R., Rosati, P., Burg, R., Mader, J., Schuecker, P., Simiç, D., Komossa, S., Reiprich, T. H., Retzlaff, J., and Trümper, J. (2000). The Northern ROSAT All-Sky (NORAS) Galaxy Cluster Survey. I. X-Ray Properties of Clusters Detected as Extended X-Ray Sources. ApJS, 129(2):435–474.

Böhringer, H. and Werner, N. (2010). X-ray spectroscopy of galaxy clusters: studying astrophysical processes in the largest celestial laboratories. A&A Rev., 18(1-2):127–196.

Bolliet, B., Brinckmann, T., Chluba, J., and Lesgourgues, J. (2019). Including massive neutrinos in thermal Sunyaev Zeldovich power spectrum and cluster counts analyses. *arXiv e-prints*, page arXiv:1906.10359.

Bond, J. R., Cole, S., Efstathiou, G., and Kaiser, N. (1991). Excursion Set Mass Functions for Hierarchical Gaussian Fluctuations. ApJ, 379:440.

Bond, J. R. and Efstathiou, G. (1984). Cosmic background radiation anisotropies in universes dominated by nonbaryonic dark matter. ApJ, 285:L45–L48.

Bond, J. R., Kofman, L., and Pogosyan, D. (1996). How filaments of galaxies are woven into the cosmic web. Nature, 380(6575):603–606.

Borgani, S., Rosati, P., Tozzi, P., Stanford, S. A., Eisenhardt, P. R., Lidman, C., Holden, B., Della Ceca, R., Norman, C., and Squires, G. (2001). Measuring Ω_m with the ROSAT Deep Cluster Survey. ApJ, 561(1):13–21.

Borm, K., Reiprich, T. H., Mohammed, I., and Lovisari, L. (2014). Constraining galaxy cluster temperatures and redshifts with eROSITA survey data. A&A, 567:A65.

Bryan, G. L. and Norman, M. L. (1998). Statistical Properties of X-Ray Clusters: Analytic and Numerical Comparisons. ApJ, 495(1):80–99.

Buote, D. A. and Humphrey, P. J. (2012). Spherically averaging ellipsoidal galaxy clusters in X-ray and Sunyaev-Zel'dovich studies - II. Biases. MNRAS, 421(2):1399–1420.

Cai, Y.-C., Cole, S., Jenkins, A., and Frenk, C. S. (2010). Full-sky map of the ISW and Rees-Sciama effect from Gpc simulations. MNRAS, 407(1):201–224.

Cai, Y.-C., Neyrinck, M., Mao, Q., Peacock, J. A., Szapudi, I., and Berlind, A. A. (2017). The lensing and temperature imprints of voids on the cosmic microwave background. MNRAS, 466:3364–3375.

Carlberg, R. G., Yee, H. K. C., and Ellingson, E. (1997). The Average Mass and Light Profiles of Galaxy Clusters. ApJ, 478(2):462–475.

Carlstrom, J. E. et al. (2011). The 10 meter south pole telescope. PASP, 123(903):568.

Carlstrom, J. E., Holder, G. P., and Reese, E. D. (2002). Cosmology with the sunyaev-zel'dovich effect. *Annual Review of Astronomy and Astrophysics*, 40(1):643–680.

Carroll, S. M. (2001). The Cosmological Constant. *Living Reviews in Relativity*, 4(1):1.

Carron, J. and Lewis, A. (2017). Maximum a posteriori CMB lensing reconstruction. Phys. Rev. D, 96(6):063510.

Carron, J., Lewis, A., and Challinor, A. (2017). Internal delensing of planck cmb temperature and polarization. *Journal of Cosmology and Astroparticle Physics*, 2017(05):035–035.

Cavaliere, A. G., Gursky, H., and Tucker, W. H. (1971). Extragalactic X-ray Sources and Associations of Galaxies. Nature, 231(5303):437–438.

Challinor, A. (2004). Anisotropies in the cosmic microwave background.

Challinor, A. and Lasenby, A. (1998). Relativistic Corrections to the Sunyaev-Zeldovich Effect. ApJ, 499(1):1–6.

Chandrasekhar, S. (1943). Stochastic Problems in Physics and Astronomy. *Reviews of Modern Physics*, 15(1):1–89.

Clowe, D., Bradač, M., Gonzalez, A. H., Markevitch, M., Randall, S. W., Jones, C., and Zaritsky, D. (2006). A Direct Empirical Proof of the Existence of Dark Matter. ApJ, 648(2):L109–L113.

Cole, S. and Efstathiou, G. (1989). Gravitational lensing of fluctuations in the microwave background radiation. MNRAS, 239:195–200.

Cole, S., Lacey, C. G., Baugh, C. M., and Frenk, C. S. (2000). Hierarchical galaxy formation. MNRAS, 319(1):168–204.

Cole, S., Percival, W. J., Peacock, J. A., Norberg, P., et al. (2005). The 2dF Galaxy Redshift Survey: power-spectrum analysis of the final data set and cosmological implications. MNRAS, 362(2):505–534.

Costanzi, M., Rozo, E., Simet, M., Zhang, Y., Evrard, A. E., Mantz, A., Rykoff, E. S., Jeltema, T., Gruen, D., Allen, S., McClintock, T., Romer, A. K., von der Linden, A., Farahi, A., DeRose, J., Varga, T. N., Weller, J., Giles, P., Hollowood, D. L., Bhargava, S., Bermeo-Hernandez, A., Chen, X., et al. (2019). Methods for cluster cosmology and application to the SDSS in preparation for DES Year 1 release. MNRAS, 488(4):4779–4800.

Dalal, N., Doré, O., Huterer, D., and Shirokov, A. (2008). Imprints of primordial non-Gaussianities on large-scale structure: Scale-dependent bias and abundance of virialized objects. Phys. Rev. D, 77(12):123514.

Das, S., Sherwin, B. D., et al. (2011). Detection of the Power Spectrum of Cosmic Microwave Background Lensing by the Atacama Cosmology Telescope. Phys. Rev. Lett., 107(2):021301.

Davis, M., Efstathiou, G., Frenk, C. S., and White, S. D. M. (1985). The evolution of large-scale structure in a universe dominated by cold dark matter. ApJ, 292:371–394.

de Bernardis, P. et al. (2000). A flat Universe from high-resolution maps of the cosmic microwave background radiation. Nature, 404(6781):955–959.

de Graaff, A., Cai, Y.-C., Heymans, C., and Peacock, J. A. (2019). Probing the missing baryons with the Sunyaev-Zel'dovich effect from filaments. A&A, 624:A48.

de Haan, T., Benson, B. A., Bleem, L. E., et al. (2016). Cosmological Constraints from Galaxy Clusters in the 2500 Square-degree SPT-SZ Survey. ApJ, 832:95.

Diaferio, A. and Geller, M. J. (1997). Infall Regions of Galaxy Clusters. ApJ, 481(2):633–643.

Dicke, R. H., Peebles, P. J. E., Roll, P. G., and Wilkinson, D. T. (1965). Cosmic Black-Body Radiation. ApJ, 142:414–419.

Diemand, J., Kuhlen, M., and Madau, P. (2007). Formation and Evolution of Galaxy Dark Matter Halos and Their Substructure. ApJ, 667(2):859–877.

Diemer, B. and Kravtsov, A. V. (2015). A Universal Model for Halo Concentrations. ApJ, 799(1):108.

Dodelson, S. (2004). CMB-cluster lensing. Phys. Rev. D, 70(2):023009.

Dodelson, S., Heitmann, K., Hirata, C., Honscheid, K., Roodman, A., Seljak, U., Slosar, A., and Trodden, M. (2016). Cosmic visions dark energy: Science.

Doroshkevich, A. G., Zel'dovich, Y. B., and Syunyaev, R. A. (1978). Fluctuations of the microwave background radiation in the adiabatic and entropic theories of galaxy formation. Soviet Ast., 22:523–528.

Dreyer, J. L. E. (1888). A New General Catalogue of Nebulæ and Clusters of Stars, being the Catalogue of the late Sir John F. W. Herschel, Bart, revised, corrected, and enlarged. MmRAS, 49:1.

Ebeling, H., Edge, A. C., Bohringer, H., Allen, S. W., Crawford, C. S., Fabian, A. C., Voges, W., and Huchra, J. P. (1998). The ROSAT Brightest Cluster Sample - I. The compilation of the sample and the cluster log N-log S distribution. MNRAS, 301(4):881–914.

Einasto, J. (1965). On the Construction of a Composite Model for the Galaxy and on the Determination of the System of Galactic Parameters. *Trudy Astrofizicheskogo Instituta Alma-Ata*, 5:87–100.

Einstein, A. (1915). Die Feldgleichungen der Gravitation. *Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften (Berlin*, pages 844–847.

Einstein, A. (1916). Die Grundlage der allgemeinen Relativitätstheorie. *Annalen der Physik*, 354(7):769–822.

Eisenstein, D. J., Zehavi, I., Hogg, D. W., Scoccimarro, R., Blanton, M. R., Nichol, R. C., Scranton, R., Seo, H.-J., Tegmark, M., Zheng, Z., et al. (2005). Detection of the Baryon Acoustic Peak in the Large-Scale Correlation Function of SDSS Luminous Red Galaxies. ApJ, 633(2):560–574.

Ettori, S., Donnarumma, A., Pointecouteau, E., Reiprich, T. H., Giodini, S., Lovisari, L., and Schmidt, R. W. (2013). Mass Profiles of Galaxy Clusters from X-ray Analysis. Space Sci. Rev., 177(1-4):119–154.

Evrard, A. E. (1988). Beyond N-body: 3D cosmological gas dynamics. MNRAS, 235:911–934.

Fabbian, G., Lewis, A., and Beck, D. (2019). CMB lensing reconstruction biases in cross-correlation with large-scale structure probes. J. Cosmology Astropart. Phys., 2019(10):057.

Fixsen, D. J. (2009). The Temperature of the Cosmic Microwave Background. ApJ, 707(2):916–920.

Friedmann, A. (1922). Über die Krümmung des Raumes. *Zeitschrift fur Physik*, 10:377–386.

Friedmann, A. (1924). Über die Möglichkeit einer Welt mit konstanter negativer Krümmung des Raumes. *Zeitschrift fur Physik*, 21(1):326–332.

Fu, L., Semboloni, E., Hoekstra, H., Kilbinger, M., van Waerbeke, L., Tereno, I., Mellier, Y., Heymans, C., Coupon, J., et al. (2008). Very weak lensing in the CFHTLS wide: cosmology from cosmic shear in the linear regime. A&A, 479(1):9–25.

Gao, L., Navarro, J. F., Frenk, C. S., Jenkins, A., Springel, V., and White, S. D. M. (2012). The Phoenix Project: the dark side of rich Galaxy clusters. MNRAS, 425(3):2169–2186.

Geach, J. E. and Peacock, J. A. (2017). Cluster richness-mass calibration with cosmic microwave background lensing. *Nature Astronomy*, 1:795–799.

Giocoli, C., Jullo, E., Metcalf, R. B., de la Torre, S., Yepes, G., Prada, F., Comparat, J., Göttlober, S., Kyplin, A., Kneib, J.-P., Petkova, M., Shan, H. Y., and Tessore, N. (2016). Multi Dark Lens Simulations: weak lensing light-cones and data base presentation. MNRAS, 461:209–223.

Górski, K. M., Hivon, E., Banday, A. J., Wandelt, B. D., Hansen, F. K., Reinecke, M., and Bartelmann, M. (2005). HEALPix: A Framework for High-Resolution Discretization and Fast Analysis of Data Distributed on the Sphere. ApJ, 622:759–771.

Gott, J. R., I. and Rees, M. J. (1975). A theory of galaxy formation and clustering. A&A, 45:365–376.

Gruen, D., Seitz, S., Becker, M. R., Friedrich, O., and Mana, A. (2015). Cosmic variance of the galaxy cluster weak lensing signal. MNRAS, 449(4):4264–4276.

Gruetjen, H. F., Fergusson, J. R., Liguori, M., and Shellard, E. P. S. (2017). Using inpainting to construct accurate cut-sky CMB estimators. Phys. Rev. D, 95(4):043532.

Gunn, J. E. and Gott, J. Richard, I. (1972). On the Infall of Matter Into Clusters of Galaxies and Some Effects on Their Evolution. ApJ, 176:1.

Gupta, N., Saro, A., Mohr, J. J., Dolag, K., and Liu, J. (2017). SZE observables, pressure profiles and centre offsets in Magneticum simulation galaxy clusters. MNRAS, 469:3069–3087.

Gursky, H., Kellogg, E., Murray, S., Leong, C., Tananbaum, H., and Giacconi, R. (1971). A Strong X-Ray Source in the Coma Cluster Observed by UHURU. ApJ, 167:L81.

Guth, A. H. (1981). Inflationary universe: A possible solution to the horizon and flatness problems. Phys. Rev. D, 23(2):347–356.

Haehnelt, M. G. and Tegmark, M. (1996). Using the Kinematic Sunyaev-Zeldovich effect to determine the peculiar velocities of clusters of galaxies. MNRAS, 279:545.

Hamana, T., Miyazaki, S., Okura, Y., Okamura, T., and Futamase, T. (2013). Toward Understanding the Anisotropic Point Spread Function of Suprime-Cam and Its Impact on Cosmic Shear Measurement. PASJ, 65:104.

Hanany, S. et al. (2000). MAXIMA-1: A Measurement of the Cosmic Microwave Background Anisotropy on Angular Scales of 10'-5°. ApJ, 545(1):L5–L9.

Hand, N. et al. (2012). Evidence of Galaxy Cluster Motions with the Kinematic Sunyaev-Zel'dovich Effect. Phys. Rev. Lett., 109(4):041101.

Hanson, D., Challinor, A., Efstathiou, G., and Bielewicz, P. (2011). CMB temperature lensing power reconstruction. Phys. Rev. D, 83(4):043005.

Hanson, D., Challinor, A., and Lewis, A. (2010). Weak lensing of the CMB. *General Relativity and Gravitation*, 42:2197–2218.

Hanson, D., Hoover, S., Crites, A., et al. (2013). Detection of B-Mode Polarization in the Cosmic Microwave Background with Data from the South Pole Telescope. Phys. Rev. Lett., 111(14):141301.

Hasselfield, M., Hilton, M., Marriage, T. A., et al. (2013). The Atacama Cosmology Telescope: Sunyaev-Zel'dovich selected galaxy clusters at 148 GHz from three seasons of data. J. Cosmology Astropart. Phys., 7:008.

He, S., Alam, S., Ferraro, S., Chen, Y.-C., and Ho, S. (2018). The detection of the imprint of filaments on cosmic microwave background lensing. *Nature Astronomy*, 2:401–406.

Heath, D. J. (1977). The growth of density perturbations in zero pressure Friedmann-Lemaître universes. MNRAS, 179:351–358.

Henden, N. A., Puchwein, E., and Sijacki, D. (2019). The redshift evolution of X-ray and Sunyaev-Zel'dovich scaling relations in the FABLE simulations. *Monthly Notices of the Royal Astronomical Society*, page 2230.

Henry, J. P. (2004). X-Ray Temperatures for the Extended Medium-Sensitivity Survey High-Redshift Cluster Sample: Constraints on Cosmology and the Dark Energy Equation of State. ApJ, 609(2):603–616.

Henry, J. P., Evrard, A. E., Hoekstra, H., Babul, A., and Mahdavi, A. (2009). The X-Ray Cluster Normalization of the Matter Power Spectrum. ApJ, 691(2):1307–1321.

Herranz, D., Sanz, J. L., Hobson, M. P., Barreiro, R. B., Diego, J. M., Martínez-González, E., and Lasenby, A. N. (2002). Filtering techniques for the detection of Sunyaev-Zel'dovich clusters <br/> in multifrequency maps. MNRAS, 336(4):1057–1068.

Herschel, W. (1785). On the Construction of the Heavens. *Philosophical Transactions of the Royal Society of London Series I*, 75:213–266.

Hill, J. C. and Spergel, D. N. (2014). Detection of thermal SZ-CMB lensing cross-correlation in Planck nominal mission data. J. Cosmology Astropart. Phys., 2014(2):030.

Hilton, M., Hasselfield, M., Sifón, C., Battaglia, N., et al. (2018). The Atacama Cosmology Telescope: The Two-season ACTPol Sunyaev-Zel'dovich Effect Selected Cluster Catalog. ApJS, 235(1):20.

Hinshaw, G., Larson, D., Komatsu, E., Spergel, D. N., et al. (2013). Nine-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Parameter Results. ApJS, 208(2):19.

Hirata, C. M., Ho, S., Padmanabhan, N., Seljak, U., and Bahcall, N. A. (2008). Correlation of CMB with large-scale structure. II. Weak lensing. Phys. Rev. D, 78(4):043520.

Hirata, C. M. and Seljak, U. (2003). Analyzing weak lensing of the cosmic microwave background using the likelihood function. Phys. Rev. D, 67(4):043001.

Hoekstra, H., Herbonnet, R., Muzzin, A., Babul, A., Mahdavi, A., Viola, M., and Cacciato, M. (2015). The Canadian Cluster Comparison Project: detailed study of systematics and updated weak lensing masses. MNRAS, 449:685–714.

Hoekstra, H., Mahdavi, A., Babul, A., and Bildfell, C. (2012). The Canadian Cluster Comparison Project: weak lensing masses and SZ scaling relations. MNRAS, 427(2):1298–1311.

Holtzman, J. A. (1989). Microwave Background Anisotropies and Large-Scale Structure in Universes with Cold Dark Matter, Baryons, Radiation, and Massive and Massless Neutrinos. ApJS, 71:1.

Horowitz, B., Ferraro, S., and Sherwin, B. D. (2019). Reconstructing small-scale lenses from the cosmic microwave background temperature fluctuations. MNRAS, 485:3919–3929.

Hu, W. (2001). Mapping the Dark Matter through the Cosmic Microwave Background Damping Tail. ApJ, 557:L79–L83.

Hu, W., DeDeo, S., and Vale, C. (2007). Cluster mass estimators from CMB temperature and polarization lensing. *New Journal of Physics*, 9:441.

Hu, W. and Dodelson, S. (2002). Cosmic Microwave Background Anisotropies. ARA&A, 40:171–216.

Hu, W. and Dodelson, S. (2002). Cosmic microwave background anisotropies. *Annual Review of Astronomy and Astrophysics*, 40(1):171–216.

Hu, W. and Kravtsov, A. (2014). massconvert: Halo Mass Conversion.

Hu, W. and Kravtsov, A. V. (2003). Sample Variance Considerations for Cluster Surveys. ApJ, 584(2):702–715.

Hu, W. and Okamoto, T. (2002). Mass Reconstruction with Cosmic Microwave Background Polarization. ApJ, 574:566–574.

Hu, W. and Sugiyama, N. (1995). Anisotropies in the Cosmic Microwave Background: an Analytic Approach. ApJ, 444:489.

Hubble, E. (1929a). A Relation between Distance and Radial Velocity among Extra-Galactic Nebulae. *Proceedings of the National Academy of Science*, 15(3):168–173.

Hubble, E. P. (1929b). A spiral nebula as a stellar system, Messier 31. ApJ, 69:103–158.

Hurier, G. and Lacasa, F. (2017). Combined analysis of galaxy cluster number count, thermal Sunyaev-Zel'dovich power spectrum, and bispectrum. A&A, 604:A71.

Hurier, G., Singh, P., and Hernández-Monteagudo, C. (2019). First measurement of the cross-correlation between CMB weak lensing and X-ray emission. A&A, 625:L4.

Huterer, D. and Shafer, D. L. (2018). Dark energy two decades after: observables, probes, consistency tests. *Reports on Progress in Physics*, 81(1):016901.

Hwang, H. S. and Lee, M. G. (2007). Searching for Rotating Galaxy Clusters in SDSS and 2dFGRS. ApJ, 662(1):236–249.

Icke, V. (1973). Formation of Galaxies inside Clusters. A&A, 27:1.

Ishak, M. (2018). Testing general relativity in cosmology. *Living Reviews in Relativity*, 22(1):1.

Itoh, N., Kohyama, Y., and Nozawa, S. (1998). Relativistic Corrections to the Sunyaev-Zeldovich Effect for Clusters of Galaxies. ApJ, 502(1):7–15.

Ivezic, Z. et al. (2008). Large Synoptic Survey Telescope: From Science Drivers To Reference Design. *Serbian Astronomical Journal*, 176:1–13.

Jaynes, E. T. (2003). *Probability theory: The logic of science*. Cambridge University Press, Cambridge.

Jenkins, A., Frenk, C. S., White, S. D. M., Colberg, J. M., Cole, S., Evrard, A. E., Couchman, H. M. P., and Yoshida, N. (2001). The mass function of dark matter haloes. MNRAS, 321(2):372–384.

Kaiser, N. (1984). On the spatial correlations of Abell clusters. ApJ, 284:L9–L12.

Kaiser, N. (1986). Evolution and clustering of rich clusters. MNRAS, 222:323–345.

Kamionkowski, M., Kosowsky, A., and Stebbins, A. (1997). A Probe of Primordial Gravity Waves and Vorticity. Phys. Rev. Lett., 78(11):2058–2061.

Kamionkowski, M. and Loeb, A. (1997). Getting around cosmic variance. Phys. Rev. D, 56(8):4511–4513.

Kant, I. (1755). *Allgemeine Naturgeschichte und Theorie des Himmels*.

Kasun, S. F. and Evrard, A. E. (2005). Shapes and Alignments of Galaxy Cluster Halos. ApJ, 629(2):781–790.

Kesden, M., Cooray, A., and Kamionkowski, M. (2003). Lensing reconstruction with CMB temperature and polarization. Phys. Rev. D, 67(12):123507.

Kirby, M., Rozo, E., Morris, R. G., Allen, S. W., Costanzi, M., Mantz, A. B., Rykoff, E. S., and von der Linden, A. (2019). Improved cosmological constraints from sdss redmapper clusters via x-ray follow-up of a complete subsample of systems.

Klypin, A. and Holtzman, J. (1997). Particle-Mesh code for cosmological simulations. *arXiv Astrophysics e-prints*.

Klypin, A., Yepes, G., Gottlöber, S., Prada, F., and Heß, S. (2016). MultiDark simulations: the story of dark matter halo concentrations and density profiles. MNRAS, 457:4340–4359.

Klypin, A. A. and Shandarin, S. F. (1983). Three-dimensional numerical model of the formation of large-scale structure in the Universe. MNRAS, 204:891–907.

Klypin, A. A., Trujillo-Gomez, S., and Primack, J. (2011). Dark Matter Halos in the Standard Cosmological Model: Results from the Bolshoi Simulation. ApJ, 740(2):102.

Knox, L. and Millea, M. (2019). The Hubble Hunter's Guide. *arXiv e-prints*, page arXiv:1908.03663.

Kodama, H. and Sasaki, M. (1984). Cosmological Perturbation Theory. *Progress of Theoretical Physics Supplement*, 78:1.

Koester, B. P., McKay, T. A., Annis, J., Wechsler, R. H., Evrard, A., Bleem, L., Becker, M., Johnston, D., Sheldon, E., Nichol, R., Miller, C., Scranton, R., et al. (2007). A MaxBCG Catalog of 13,823 Galaxy Clusters from the Sloan Digital Sky Survey. ApJ, 660(1):239–255.

Komatsu, E. and Seljak, U. (2002). The Sunyaev-Zel'dovich angular power spectrum as a probe of cosmological parameters. MNRAS, 336(4):1256–1270.

Kosowsky, A., Milosavljevic, M., and Jimenez, R. (2002). Efficient cosmological parameter estimation from microwave background anisotropies. Phys. Rev. D, 66(6):063007.

Kovac, J. M., Leitch, E. M., Pryke, C., Carlstrom, J. E., Halverson, N. W., and Holzapfel, W. L. (2002). Detection of polarization in the cosmic microwave background using DASI. Nature, 420(6917):772–787.

Kravtsov, A. V. and Borgani, S. (2012). Formation of Galaxy Clusters. ARA&A, 50:353–409.

Kravtsov, A. V., Vikhlinin, A., and Nagai, D. (2006). A New Robust Low-Scatter X-Ray Mass Indicator for Clusters of Galaxies. ApJ, 650(1):128–136.

Laplace, P. S. (1774). Mémoire sur la probabilité des causes par les événemens. *Mem. Acad. Roy. Sci.*, 6:621–656.

Lau, E. T., Kravtsov, A. V., and Nagai, D. (2009). Residual Gas Motions in the Intracluster Medium and Bias in Hydrostatic Measurements of Mass Profiles of Clusters. ApJ, 705(2):1129–1138.

Le Brun, A. M. C., McCarthy, I. G., Schaye, J., and Ponman, T. J. (2014). Towards a realistic population of simulated galaxy groups and clusters. MNRAS, 441(2):1270–1290.

Lemaître, G. (1927). Un Univers homogène de masse constante et de rayon croissant rendant compte de la vitesse radiale des nébuleuses extra-galactiques. *Annales de la Société Scientifique de Bruxelles*, 47:49–59.

Lewis, A. and Challinor, A. (2006). Weak gravitational lensing of the CMB. Phys. Rep., 429:1–65.

Lewis, A., Challinor, A., and Hanson, D. (2011). The shape of the CMB lensing bispectrum. J. Cosmology Astropart. Phys., 3:018.

Lewis, A. and King, L. (2006). Cluster masses from CMB and galaxy weak lensing. Phys. Rev. D, 73(6):063006.

Liddle, A. R. and Lyth, D. H. (2000). *Cosmological Inflation and Large-Scale Structure*.

Lifshitz, E. M. (1946). On the gravitational stability of the expanding universe. *Zhurnal Eksperimentalnoi i Teoreticheskoi Fiziki*, 16:587–602.

Limber, D. N. (1959). Effects of Intracluster Gas and Duct upon the Virial Theorem. ApJ, 130:414.

Linde, A. D. (1982). A new inflationary universe scenario: A possible solution of the horizon, flatness, homogeneity, isotropy and primordial monopole problems. *Physics Letters B*, 108(6):389–393.

Linde, A. D. (1983). Chaotic inflation. *Physics Letters B*, 129(3-4):177–181.

Louis, T. and Alonso, D. (2017). Calibrating cluster number counts with CMB lensing. Phys. Rev. D, 95(4):043517.

Lynden-Bell, D. (1967). Statistical mechanics of violent relaxation in stellar systems. MNRAS, 136:101.

Mackay, D. J. C. (2003). *Information Theory, Inference and Learning Algorithms*.

Madhavacheril, M., Sehgal, N., et al. (2015). Evidence of Lensing of the Cosmic Microwave Background by Dark Matter Halos. Phys. Rev. Lett., 114(15):151302.

Mantz, A., Allen, S. W., Ebeling, H., and Rapetti, D. (2008). New constraints on dark energy from the observed growth of the most X-ray luminous galaxy clusters. MNRAS, 387(3):1179–1192.

Mantz, A., Allen, S. W., Rapetti, D., and Ebeling, H. (2010). The observed growth of massive galaxy clusters - I. Statistical methods and cosmological constraints. MNRAS, 406:1759–1772.

Mantz, A. B., Abdulla, Z., Allen, S. W., Carlstrom, J. E., Logan, C. H. A., Marrone, D. P., Maughan, B. J., Willis, J., Pacaud, F., and Pierre, M. (2018). The XXL Survey. XVII. X-ray and Sunyaev-Zel'dovich properties of the redshift 2.0 galaxy cluster XLSSC 122. A&A, 620:A2.

Mantz, A. B., von der Linden, A., et al. (2015). Weighing the giants - IV. Cosmology and neutrino mass. MNRAS, 446(3):2205–2225.

Markevitch, M. (2006). Chandra Observation of the Most Interesting Cluster in the Universe. In Wilson, A., editor, *The X-ray Universe 2005*, volume 604 of *ESA Special Publication*, page 723.

Markevitch, M., Gonzalez, A. H., Clowe, D., Vikhlinin, A., Forman, W., Jones, C., Murray, S., and Tucker, W. (2004). Direct Constraints on the Dark Matter Self-Interaction Cross Section from the Merging Galaxy Cluster 1E 0657-56. ApJ, 606(2):819–824.

Mather, J. C. et al. (1994). Measurement of the Cosmic Microwave Background Spectrum by the COBE FIRAS Instrument. ApJ, 420:439.

McClintock, T., Varga, T. N., Gruen, D., Rozo, E., Rykoff, E. S., Shin, T., Melchior, P., DeRose, J., Seitz, S., Dietrich, J. P., Sheldon, E., Zhang, Y., von der Linden, A., Jeltema, T., Mantz, A. B., Romer, A. K., et al. (2019). Dark Energy Survey Year 1 results: weak lensing mass calibration of redMaPPer galaxy clusters. MNRAS, 482(1):1352–1378.

Medezinski, E., Battaglia, N., Umetsu, K., Oguri, M., Miyatake, H., Nishizawa, A. J., Sifón, C., Spergel, D. N., Chiu, I.-N., Lin, Y.-T., Bahcall, N., and Komiyama, Y. (2018). Planck Sunyaev-Zel'dovich cluster mass calibration using Hyper Suprime-Cam weak lensing. PASJ, 70:S28.

Meekins, J. F., Fritz, G., Chubb, T. A., and Friedman, H. (1971). Physical Sciences: X-rays from the Coma Cluster of Galaxies. Nature, 231(5298):107–108.

Melin, J.-B. and Bartlett, J. G. (2015). Measuring cluster masses with CMB lensing: a statistical approach. A&A, 578:A21.

Melin, J. B., Bartlett, J. G., and Delabrouille, J. (2006). Catalog extraction in SZ cluster surveys: a matched filter approach. A&A, 459(2):341–352.

Meneghetti, M., Rasia, E., Merten, J., Bellagamba, F., Ettori, S., Mazzotta, P., Dolag, K., and Marri, S. (2010). Weighing simulated galaxy clusters using lensing and X-ray. A&A, 514:A93.

Messier, C. (1784). *La Connoissance des temps*.

Meszaros, P. (1974). The behaviour of point masses in an expanding cosmological substratum. A&A, 37(2):225–228.

Millea, M., Anderes, E., and Wandelt, B. D. (2019). Bayesian delensing of CMB temperature and polarization. Phys. Rev. D, 100(2):023509.

Miller, A. D., Caldwell, R., Devlin, M. J., Dorwart, W. B., Herbig, T., Nolta, M. R., Page, L. A., Puchalla, J., Torbet, E., and Tran, H. T. (1999). A Measurement of the Angular Power Spectrum of the Cosmic Microwave Background from L = 100 to 400. ApJ, 524(1):L1–L4.

Miyatake, H., Battaglia, N., Hilton, M., Medezinski, E., Nishizawa, A. J., More, S., et al. (2019). Weak-lensing Mass Calibration of ACTPol Sunyaev-Zel'dovich Clusters with the Hyper Suprime-Cam Survey. ApJ, 875:63.

Miyazaki, S., Oguri, M., Hamana, T., Shirasaki, M., Koike, M., Komiyama, Y., Umetsu, K., Utsumi, Y., Okabe, N., More, S., Medezinski, E., Lin, Y.-T., Miyatake, H., Murayama, H., Ota, N., and Mitsuishi, I. (2018). A large sample of shear-selected clusters from the Hyper Suprime-Cam Subaru Strategic Program S16A Wide field mass maps. PASJ, 70:S27.

Mortonson, M. J., Hu, W., and Huterer, D. (2011). Simultaneous falsification of ΛCDM and quintessence with massive, distant clusters. Phys. Rev. D, 83(2):023015.

Mroczkowski, T. et al. (2009). Application of a Self-Similar Pressure Profile to Sunyaev-Zel'dovich Effect Data from Galaxy Clusters. ApJ, 694(2):1034–1044.

Mroczkowski, T., Nagai, D., et al. (2019). Astrophysics with the Spatially and Spectrally Resolved Sunyaev-Zeldovich Effects. A Millimetre/Submillimetre Probe of the Warm and Hot Universe. Space Sci. Rev., 215(1):17.

Muñoz-Cuartas, J. C., Macciò, A. V., Gottlöber, S., and Dutton, A. A. (2011). The redshift evolution of Λ cold dark matter halo parameters: concentration, spin and shape. MNRAS, 411:584–594.

Mukhanov, V. F. and Chibisov, G. V. (1981). Quantum fluctuations and a nonsingular universe. *Soviet Journal of Experimental and Theoretical Physics Letters*, 33:532.

Mukhanov, V. F., Feldman, H. A., and Brandenberger, R. H. (1992). Theory of cosmological perturbations. Phys. Rep., 215(5-6):203–333.

Nagai, D., Kravtsov, A. V., and Vikhlinin, A. (2007a). Effects of Galaxy Formation on Thermodynamics of the Intracluster Medium. ApJ, 668:1–14.

Nagai, D., Vikhlinin, A., and Kravtsov, A. V. (2007b). Testing X-Ray Measurements of Galaxy Clusters with Cosmological Simulations. ApJ, 655(1):98–108.

Nagarajan, A., Pacaud, F., Sommer, M., Klein, M., et al. (2019). Weak-lensing mass calibration of the Sunyaev-Zel'dovich effect using APEX-SZ galaxy clusters. MNRAS, 488:1728–1759.

Namikawa, T., Hanson, D., and Takahashi, R. (2013). Bias-hardened CMB lensing. MNRAS, 431:609–620.

Navarro, J. F., Frenk, C. S., and White, S. D. M. (1995). Simulations of X-ray clusters. MNRAS, 275(3):720–740.

Navarro, J. F., Frenk, C. S., and White, S. D. M. (1997). A Universal Density Profile from Hierarchical Clustering. ApJ, 490:493–508.

Nelson, K., Lau, E. T., and Nagai, D. (2014a). Hydrodynamic Simulation of Non-thermal Pressure Profiles of Galaxy Clusters. ApJ, 792(1):25.

Nelson, K., Lau, E. T., Nagai, D., Rudd, D. H., and Yu, L. (2014b). Weighing Galaxy Clusters with Gas. II. On the Origin of Hydrostatic Mass Bias in ΛCDM Galaxy Clusters. ApJ, 782(2):107.

Newton, I. (1687). *Philosophiae Naturalis Principia Mathematica. Auctore Js. Newton.*

O'Brien, T. A., Collins, W. D., Rauscher, S. A., and Ringler, T. D. (2014). Reducing the computational cost of the ecf using a nufft: A fast and objective probability density estimation method. *Computational Statistics & Data Analysis*, 79:222 – 234.

O'Brien, T. A., Kashinath, K., Cavanaugh, N. R., Collins, W. D., and O'Brien, J. P. (2016). A fast and objective multidimensional kernel density estimation method: fastkde. *Computational Statistics & Data Analysis*, 101:148 – 160.

Oguri, M. and Hamana, T. (2011). Detailed cluster lensing profiles at large radii and the impact on cluster weak lensing studies. MNRAS, 414:1851–1861.

Okabe, N. and Smith, G. P. (2016). LoCuSS: weak-lensing mass calibration of galaxy clusters. MNRAS, 461(4):3794–3821.

Okabe, N., Takada, M., Umetsu, K., Futamase, T., and Smith, G. P. (2010). LoCuSS: Subaru Weak Lensing Study of 30 Galaxy Clusters. PASJ, 62:811.

Okamoto, T. and Hu, W. (2003). Cosmic microwave background lensing reconstruction on the full sky. Phys. Rev. D, 67(8):083002.

Ota, N., Nagai, D., and Lau, E. T. (2018). Constraining hydrostatic mass bias of galaxy clusters with high-resolution X-ray spectroscopy. PASJ, 70(3):51.

Pacaud, F., Pierre, M., Melin, J. B., et al. (2018). The XXL Survey. XXV. Cosmological analysis of the C1 cluster number counts. A&A, 620:A10.

Patil, S., Raghunathan, S., and Reichardt, C. L. (2020). Suppressing the Thermal SZ-induced Variance in CMB-cluster Lensing Estimators. ApJ, 888(1):9.

Peacock, J. A. (1999). *Cosmological Physics.*

Peebles, P. J. E. (1968). Recombination of the Primeval Plasma. ApJ, 153:1.

Peebles, P. J. E. (1980). *The large-scale structure of the universe.*

Peebles, P. J. E. and Yu, J. T. (1970). Primeval Adiabatic Perturbation in an Expanding Universe. ApJ, 162:815.

Penna-Lima, M., Bartlett, J. G., Rozo, E., Melin, J.-B., Merten, J., Evrard, A. E., Postman, M., and Rykoff, E. (2017). Calibrating the Planck cluster mass scale with CLASH. A&A, 604:A89.

Penzias, A. A. and Wilson, R. W. (1965). A Measurement of Excess Antenna Temperature at 4080 Mc/s. ApJ, 142:419–421.

Percival, W. J. (2005). Cosmological structure formation in a homogeneous dark energy background. A&A, 443(3):819–830.

Perlmutter, S., Aldering, G., Goldhaber, G., Knop, R. A., Nugent, P., et al. (1999). Measurements of Ω and Λ from 42 High-Redshift Supernovae. ApJ, 517(2):565–586.

Pierpaoli, E., Borgani, S., Scott, D., and White, M. (2003). On determining the cluster abundance normalization. MNRAS, 342(1):163–175.

Piffaretti, R. and Valdarnini, R. (2008). Total mass biases in X-ray galaxy clusters. A&A, 491:71–87.

Pillepich, A., Porciani, C., and Reiprich, T. H. (2012). The X-ray cluster survey with eRosita: forecasts for cosmology, cluster physics and primordial non-Gaussianity. MNRAS, 422(1):44–69.

Planck 2013 results XVIII (2014). Planck 2013 results. XVIII. Gravitational lensing-infrared background correlation. A&A, 571:A18.

Planck 2013 results XX (2014). Planck 2013 results. XX. Cosmology from Sunyaev-Zeldovich cluster counts. A&A, 571:A20.

Planck 2013 results XXIX (2014). Planck 2013 results. XXIX. The Planck catalogue of Sunyaev-Zeldovich sources. A&A, 571:A29.

Planck 2015 results VIII (2016). Planck 2015 results. VIII. High Frequency Instrument data processing: Calibration and maps. A&A, 594:A8.

Planck 2015 results XIII (2016). Planck 2015 results. XIII. Cosmological parameters. A&A, 594:A13.

Planck 2015 results XV (2016). Planck 2015 results. xv. gravitational lensing. *Astronomy & Astrophysics*, 594:A15.

Planck 2015 results XXIV (2016). Planck 2015 results. XXIV. Cosmology from Sunyaev-Zeldovich cluster counts. A&A, 594:A24.

Planck 2015 results XXVII (2016). Planck 2015 results. XXVII. The second Planck catalogue of Sunyaev-Zeldovich sources. A&A, 594:A27.

Planck 2018 results I (2018). Planck 2018 results. I. Overview and the cosmological legacy of Planck. *arXiv e-prints*, page arXiv:1807.06205.

Planck 2018 results VI (2018). Planck 2018 results. VI. Cosmological parameters. *arXiv e-prints*, page arXiv:1807.06209.

Planck 2018 results VIII (2018). Planck 2018 results. VIII. Gravitational lensing. *arXiv e-prints*, page arXiv:1807.06210.

Planck 2018 results X (2018). Planck 2018 results. X. Constraints on inflation. *arXiv e-prints*, page arXiv:1807.06211.

Planelles, S., Fabjan, D., Borgani, S., Murante, G., Rasia, E., Biffi, V., Truong, N., Ragone-Figueroa, C., Granato, G. L., Dolag, K., Pierpaoli, E., Beck, A. M., Steinborn, L. K., and Gaspari, M. (2017). Pressure of the hot gas in simulations of galaxy clusters. *Monthly Notices of the Royal Astronomical Society*, 467(4):3827–3847.

Polnarev, A. G. (1985). Polarization and Anisotropy Induced in the Microwave Background by Cosmological Gravitational Waves. Soviet Ast., 29:607–613.

Pratt, G. W., Arnaud, M., Biviano, A., Eckert, D., Ettori, S., Nagai, D., Okabe, N., and Reiprich, T. H. (2019). The Galaxy Cluster Mass Scale and Its Impact on Cosmological Constraints from the Cluster Population. Space Sci. Rev., 215:25.

Pratt, G. W., Croston, J. H., Arnaud, M., and Böhringer, H. (2009). Galaxy cluster X-ray luminosity scaling relations from a representative local sample (REXCESS). A&A, 498:361–378.

Press, W. H. and Schechter, P. (1974). Formation of Galaxies and Clusters of Galaxies by Self-Similar Gravitational Condensation. ApJ, 187:425–438.

Raghunathan, S., Holder, G. P., Bartlett, J. G., Patil, S., Reichardt, C. L., and Whitehorn, N. (2019a). An inpainting approach to tackle the kinematic and thermal SZ induced biases in CMB-cluster lensing estimators. J. Cosmology Astropart. Phys., 2019(11):037.

Raghunathan, S., Nadathur, S., Sherwin, B. D., and Whitehorn, N. (2020). The Gravitational Lensing Signatures of BOSS Voids in the Cosmic Microwave Background. ApJ, 890(2):168.

Raghunathan, S., Patil, S., Baxter, E., Benson, B. A., Bleem, L. E., Chou, T. L., Crawford, T. M., Holder, G. P., McClintock, T., Reichardt, C. L., Rozo, E., Varga, T. N., et al. (2019b). Mass Calibration of Optically Selected DES Clusters Using a Measurement of CMB-cluster Lensing with SPTpol Data. ApJ, 872(2):170.

Raghunathan, S., Patil, S., Baxter, E., Benson, B. A., Bleem, L. E., Crawford, T. M., Holder, G. P., McClintock, T., Reichardt, C. L., Varga, T. N., Whitehorn, N., et al. (2019c). A Detection of CMB-Cluster Lensing using Polarization Data from SPTpol. *arXiv e-prints*.

Raghunathan, S., Patil, S., Baxter, E. J., Bianchini, F., Bleem, L. E., Crawford, T. M., Holder, G. P., Manzotti, A., and Reichardt, C. L. (2017). Measuring galaxy cluster masses with CMB lensing using a Maximum Likelihood estimator: statistical and systematic error budgets for future experiments. J. Cosmology Astropart. Phys., 8:030.

Randall, S. W., Markevitch, M., Clowe, D., Gonzalez, A. H., and Bradač, M. (2008). Constraints on the Self-Interaction Cross Section of Dark Matter from Numerical Simulations of the Merging Galaxy Cluster 1E 0657-56. ApJ, 679(2):1173–1180.

Rasia, E., Ettori, S., Moscardini, L., Mazzotta, P., Borgani, S., Dolag, K., Tormen, G., Cheng, L. M., and Diaferio, A. (2006). Systematics in the X-ray cluster mass estimators. MNRAS, 369(4):2013–2024.

Rasia, E., Lau, E. T., Borgani, S., Nagai, D., Dolag, K., Avestruz, C., Granato, G. L., Mazzotta, P., Murante, G., Nelson, K., and Ragone-Figueroa, C. (2014). Temperature Structure of the Intracluster Medium from Smoothed-particle Hydrodynamics and Adaptive-mesh Refinement Simulations. ApJ, 791(2):96.

Rasia, E., Meneghetti, M., Martino, R., Borgani, S., Bonafede, A., Dolag, K., Ettori, S., Fabjan, D., Giocoli, C., Mazzotta, P., Merten, J., Radovich, M., and Tornatore, L. (2012). Lensing and x-ray mass estimates of clusters (simulations). *New Journal of Physics*, 14(5):055018.

Rees, M. J. (1968). Polarization and Spectrum of the Primeval Radiation in an Anisotropic Universe. ApJ, 153:L1.

Rees, M. J. and Sciama, D. W. (1968). Large-scale Density Inhomogeneities in the Universe. Nature, 217(5128):511–516.

Reiprich, T. H. and Böhringer, H. (2002). The Mass Function of an X-Ray Flux-limited Sample of Galaxy Clusters. ApJ, 567(2):716–740.

Remazeilles, M., Bolliet, B., Rotti, A., and Chluba, J. (2019). Can we neglect relativistic temperature corrections in the Planck thermal SZ analysis? MNRAS, 483(3):3459–3464.

Remazeilles, M., Delabrouille, J., and Cardoso, J.-F. (2011). CMB and SZ effect separation with constrained Internal Linear Combinations. MNRAS, 410:2481–2487.

Rephaeli, Y. (1995a). Comptonization Of The Cosmic Microwave Background: The Sunyaev-Zeldovich Effect. ARA&A, 33:541–580.

Rephaeli, Y. (1995b). Cosmic Microwave Background Comptonization by Hot Intracluster Gas. ApJ, 445:33.

Riebe, K., Partl, A. M., Enke, H., Forero-Romero, J., Gottlöber, S., Klypin, A., Lemson, G., Prada, F., Primack, J. R., Steinmetz, M., and Turchaninov, V. (2013). The MultiDark Database: Release of the Bolshoi and MultiDark cosmological simulations. *Astronomische Nachrichten*, 334:691–708.

Riess, A. G., Casertano, S., Yuan, W., Macri, L. M., and Scolnic, D. (2019). Large Magellanic Cloud Cepheid Standards Provide a 1% Foundation for the Determination of the Hubble Constant and Stronger Evidence for Physics beyond ΛCDM. ApJ, 876(1):85.

Riess, A. G., Filippenko, A. V., Challis, P., Clocchiatti, A., Diercks, A., Garnavich, P. M., Gilliland, R. L., Hogan, C. J., Jha, S., Kirshner, R. P., Leibundgut, B., Phillips, M. M., Reiss, D., Schmidt, B. P., Schommer, R. A., Smith, R. C., Spyromilio, J., Stubbs, C., Suntzeff, N. B., and Tonry, J. (1998). Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant. AJ, 116(3):1009–1038.

Robertson, H. P. (1932). The Expanding Universe. *Science*, 76:221–226.

Roncarelli, M., Ettori, S., Borgani, S., Dolag, K., Fabjan, D., and Moscardini, L. (2013). Large-scale inhomogeneities of the intracluster medium: improving mass estimates using the observed azimuthal scatter. MNRAS, 432(4):3030–3046.

Roncarelli, M., Gaspari, M., Ettori, S., Biffi, V., Brighenti, F., Bulbul, E., Clerc, N., Cucchetti, E., Pointecouteau, E., and Rasia, E. (2018). Measuring turbulence and gas motions in galaxy clusters via synthetic Athena X-IFU observations. A&A, 618:A39.

Rossetti, M., Gastaldello, F., Eckert, D., Della Torre, M., Pantiri, G., Cazzoletti, P., and Molendi, S. (2017). The cool-core state of Planck SZ-selected clusters versus X-ray-selected samples: evidence for cool-core bias. MNRAS, 468(2):1917–1930.

Rossetti, M., Gastaldello, F., Ferioli, G., Bersanelli, M., De Grandi, S., Eckert, D., Ghizzardi, S., Maino, D., and Molendi, S. (2016). Measuring the dynamical state of Planck SZ-selected clusters: X-ray peak - BCG offset. MNRAS, 457(4):4515–4524.

Rozo, E., Wechsler, R. H., Rykoff, E. S., Annis, J. T., Becker, M. R., Evrard, A. E., Frieman, J. A., Hansen, S. M., Hao, J., Johnston, D. E., Koester, B. P., McKay, T. A., Sheldon, E. S., and Weinberg, D. H. (2010). Cosmological Constraints from the Sloan Digital Sky Survey maxBCG Cluster Catalog. ApJ, 708(1):645–660.

Ruan, J. J., Quinn, T. R., and Babul, A. (2013). The observable thermal and kinetic Sunyaev-Zel'dovich effect in merging galaxy clusters. MNRAS, 432(4):3508–3519.

Ruppin, F., Mayet, F., Macías-Pérez, J. F., and Perotto, L. (2019). Impact of the mean pressure profile of galaxy clusters on the cosmological constraints from the Planck tSZ power spectrum. MNRAS, 490(1):784–796.

Sachs, R. K. and Wolfe, A. M. (1967). Perturbations of a Cosmological Model and Angular Variations of the Microwave Background. ApJ, 147:73.

Sahlén, M., Zubeldía, Í., and Silk, J. (2016). Cluster-Void Degeneracy Breaking: Dark Energy, Planck, and the Largest Cluster and Void. ApJ, 820(1):L7.

Salvati, L., Douspis, M., and Aghanim, N. (2018). Constraints from thermal Sunyaev-Zel'dovich cluster counts and power spectrum combined with CMB. A&A, 614:A13.

Sánchez, C., Carrasco Kind, M., Lin, H., Miquel, R., et al. (2014). Photometric redshift analysis in the Dark Energy Survey Science Verification data. MNRAS, 445(2):1482–1506.

Sartoris, B., Biviano, A., Fedeli, C., Bartlett, J. G., Borgani, S., Costanzi, M., Giocoli, C., Moscardini, L., Weller, J., Ascaso, B., Bardelli, S., Maurogordato, S., and Viana, P. T. P. (2016). Next generation cosmology: constraints from the Euclid galaxy cluster survey. MNRAS, 459(2):1764–1780.

Sasaki, S. (1996). A New Method to Estimate Cosmological Parameters Using the Baryon Fraction of Clusters of Galaxies. PASJ, 48:L119–L122.

Sayre, J. T., Reichardt, C. L., Henning, J. W., et al. (2019). Measurements of B-mode Polarization of the Cosmic Microwave Background from 500 Square Degrees of SPTpol Data. *arXiv e-prints*, page arXiv:1910.05748.

Schaan, E. and Ferraro, S. (2019). Foreground-Immune Cosmic Microwave Background Lensing with Shear-Only Reconstruction. Phys. Rev. Lett., 122(18):181301.

Schellenberger, G. and Reiprich, T. H. (2017). HICOSMO: cosmology with a complete sample of galaxy clusters - II. Cosmological results. MNRAS, 471(2):1370–1389.

Schellenberger, G., Reiprich, T. H., Lovisari, L., Nevalainen, J., and David, L. (2015). XMM-Newton and Chandra cross-calibration using HIFLUGCS galaxy clusters . Systematic temperature differences and cosmological impact. A&A, 575:A30.

Scherrer, R. J. and Weinberg, D. H. (1998). Constraints on the Effects of Locally Biased Galaxy Formation. ApJ, 504(2):607–611.

Schuecker, P., Böhringer, H., Collins, C. A., and Guzzo, L. (2003). The REFLEX galaxy cluster survey. VII. Omega$_m$ and sigma$_8$ from cluster abundance and large-scale clustering. A&A, 398:867–877.

Schumann, M. (2019). Direct detection of WIMP dark matter: concepts and status. *Journal of Physics G Nuclear Physics*, 46(10):103003.

Sehgal, N., Bode, P., Das, S., Hernand ez-Monteagudo, C., Huffenberger, K., Lin, Y.-T., Ostriker, J. P., and Trac, H. (2010). Simulations of the Microwave Sky. ApJ, 709(2):920–936.

Sehgal, N., Trac, H., Acquaviva, V., et al. (2011). The Atacama Cosmology Telescope: Cosmology from Galaxy Clusters Detected via the Sunyaev-Zel'dovich Effect. ApJ, 732(1):44.

Seljak, U. (1994). A Two-Fluid Approximation for Calculating the Cosmic Microwave Background Anisotropies. ApJ, 435:L87.

Seljak, U. (2002). Cluster number density normalization from the observed mass-temperature relation. MNRAS, 337(3):769–773.

Seljak, U. and Hirata, C. M. (2004). Gravitational lensing as a contaminant of the gravity wave signal in the cmb. *Physical Review D*, 69(4).

Seljak, U., Sugiyama, N., White, M., and Zaldarriaga, M. (2003). Comparison of cosmological Boltzmann codes: Are we ready for high precision cosmology? Phys. Rev. D, 68(8):083507.

Seljak, U. and Zaldarriaga, M. (1997). Signature of Gravity Waves in the Polarization of the Microwave Background. Phys. Rev. Lett., 78(11):2054–2057.

Seljak, U. and Zaldarriaga, M. (2000). Lensing-induced Cluster Signatures in the Cosmic Microwave Background. ApJ, 538:57–64.

Sereno, M., Covone, G., Izzo, L., Ettori, S., Coupon, J., and Lieu, M. (2017). PSZ2LenS. Weak lensing analysis of the Planck clusters in the CFHTLenS and in the RCSLenS. MNRAS, 472:1946–1971.

Sereno, M. and Ettori, S. (2015). Comparing masses in literature (CoMaLit) - I. Bias and scatter in weak lensing and X-ray mass estimates of clusters. MNRAS, 450:3633–3648.

Shaw, L. D., Holder, G. P., and Dudley, J. (2010). Non-Gaussian Scatter in Cluster Scaling Relations. ApJ, 716:281–285.

Sheth, R. K., Mo, H. J., and Tormen, G. (2001). Ellipsoidal collapse and an improved model for the number and spatial distribution of dark matter haloes. MNRAS, 323(1):1–12.

Sheth, R. K. and Tormen, G. (2002). An excursion set model of hierarchical clustering: ellipsoidal collapse and the moving barrier. MNRAS, 329(1):61–75.

Shi, X., Komatsu, E., Nagai, D., and Lau, E. T. (2016). Analytical model for non-thermal pressure in galaxy clusters - III. Removing the hydrostatic mass bias. MNRAS, 455(3):2936–2944.

Shirasaki, M., Nagai, D., and Lau, E. T. (2016). Covariance in the thermal SZ-weak lensing mass scaling relation of galaxy clusters. MNRAS, 460:3913–3924.

Silk, J. (1968). Cosmic Black-Body Radiation and Galaxy Formation. ApJ, 151:459.

Silk, J. and White, S. D. M. (1978). The determination of q_0 using X-ray and microwave observation of galaxy clusters. ApJ, 226:L103–L106.

Simons Observatory Collaboration (2019). The Simons Observatory: science goals and forecasts. J. Cosmology Astropart. Phys., 2019(2):056.

Smith, K. M., Zahn, O., and Doré, O. (2007). Detection of gravitational lensing in the cosmic microwave background. *Phys. Rev. D*, 76:043510.

Smith, R. E. and Marian, L. (2011). What do cluster counts really tell us about the Universe? MNRAS, 418(2):729–746.

Smoot, G. F., Bennett, C. L., Kogut, A., Wright, E. L., et al. (1992). Structure in the COBE Differential Microwave Radiometer First-Year Maps. ApJ, 396:L1.

Springel, V., White, S. D. M., Jenkins, A., Frenk, C. S., Yoshida, N., Gao, L., Navarro, J., Thacker, R., Croton, D., Helly, J., Peacock, J. A., Cole, S., Thomas, P., Couchman, H., Evrard, A., Colberg, J., and Pearce, F. (2005). Simulations of the formation, evolution and clustering of galaxies and quasars. Nature, 435(7042):629–636.

Staniszewski, Z. et al. (2009). Galaxy Clusters Discovered with a Sunyaev-Zel'dovich Effect Survey. ApJ, 701:32–41.

Stebbins, A. (1997). Extensions to the Kompaneets Equation and Sunyaev-Zel'dovich Distortion. *arXiv e-prints*, pages astro–ph/9709065.

Sunyaev, R. A. and Zeldovich, I. B. (1980). The velocity of clusters of galaxies relative to the microwave background - The possibility of its measurement. MNRAS, 190:413–420.

Sunyaev, R. A. and Zeldovich, Y. B. (1972). The Observations of Relic Radiation as a Test of the Nature of X-Ray Radiation from the Clusters of Galaxies. *Comments on Astrophysics and Space Physics*, 4:173.

Takizawa, M., Nagino, R., and Matsushita, K. (2010). Mass Estimation of Merging Galaxy Clusters. PASJ, 62:951.

Tegmark, M., Strauss, M. A., Blanton, M. R., Abazajian, K., Dodelson, S., Sandvik, H., Wang, X., Weinberg, D. H., Zehavi, I., Bahcall, N. A., Hoyle, F., Schlegel, D., Scoccimarro, R., Vogeley, M. S., et al. (2004). Cosmological parameters from SDSS and WMAP. Phys. Rev. D, 69(10):103501.

Tinker, J., Kravtsov, A. V., Klypin, A., Abazajian, K., Warren, M., Yepes, G., Gottlöber, S., and Holz, D. E. (2008). Toward a Halo Mass Function for Precision Cosmology: The Limits of Universality. ApJ, 688:709–728.

Tinker, J. L., Robertson, B. E., Kravtsov, A. V., Klypin, A., Warren, M. S., Yepes, G., and Gottlöber, S. (2010). The Large-scale Bias of Dark Matter Halos: Numerical Calibration and Model Tests. ApJ, 724:878–886.

Truong, N., Rasia, E., Mazzotta, P., Planelles, S., Biffi, V., Fabjan, D., Beck, A. M., Borgani, S., Dolag, K., Gaspari, M., Granato, G. L., Murante, G., Ragone-Figueroa, C., and Steinborn, L. K. (2017). Cosmological hydrodynamical simulations of galaxy clusters: X-ray scaling relations and their evolution. *Monthly Notices of the Royal Astronomical Society*, 474(3):4089–4111.

Umetsu, K., Medezinski, E., Nonino, M., Merten, J., Postman, M., Meneghetti, M., Donahue, M., Czakon, N., Molino, A., Seitz, S., Gruen, D., Lemze, D., et al. (2014). CLASH: Weak-lensing Shear-and-magnification Analysis of 20 Galaxy Clusters. ApJ, 795(2):163.

Vale, C., Amblard, A., and White, M. (2004). Cluster lensing of the CMB. New A, 10:1–15.

van Albada, G. B. (1960). Formation and evolution of clusters of galaxies (Errata: 15 330). Bull. Astron. Inst. Netherlands, 15:165.

van Uitert, E., Joachimi, B., Joudaki, S., Amon, A., Heymans, C., Köhlinger, F., Asgari, M., Blake, C., Choi, A., Erben, T., Farrow, D. J., Harnois-Déraps, J., Hildebrandt, H., Hoekstra, H., Kitching, T. D., Klaes, D., Kuijken, K., Merten, J., Miller, L., Nakajima, R., Schneider, P., Valentijn, E., and Viola, M. (2018). KiDS+GAMA: cosmology constraints from a joint analysis of cosmic shear, galaxy-galaxy lensing, and angular clustering. MNRAS, 476(4):4662–4689.

Vanderlinde, K., Crawford, T. M., de Haan, T., Dudley, J. P., Shaw, L., et al. (2010). Galaxy Clusters Selected with the Sunyaev-Zel'dovich Effect from 2008 South Pole Telescope Observations. ApJ, 722(2):1180–1196.

Vazza, F., Brunetti, G., Kritsuk, A., Wagner, R., Gheller, C., and Norman, M. (2009). Turbulent motions and shocks waves in galaxy clusters simulated with adaptive mesh refinement. A&A, 504(1):33–43.

Viana, P. T. P., Nichol, R. C., and Liddle, A. R. (2002). Constraining the Matter Power Spectrum Normalization Using the Sloan Digital Sky Survey/ROSAT All-Sky Survey and REFLEX Cluster Surveys. ApJ, 569(2):L75–L78.

Vikhlinin, A., Kravtsov, A. V., Burenin, R. A., Ebeling, H., Forman, W. R., Hornstrup, A., Jones, C., Murray, S. S., Nagai, D., Quintana, H., and Voevodkin, A. (2009). Chandra Cluster Cosmology Project III: Cosmological Parameter Constraints. ApJ, 692(2):1060–1074.

Voevodkin, A. and Vikhlinin, A. (2004). Constraining Amplitude and Slope of the Mass Fluctuation Spectrum Using a Cluster Baryon Mass Function. ApJ, 601(2):610–620.

Vogelsberger, M., Genel, S., Springel, V., Torrey, P., Sijacki, D., Xu, D., Snyder, G., Nelson, D., and Hernquist, L. (2014). Introducing the Illustris Project: simulating the coevolution of dark and visible matter in the Universe. MNRAS, 444(2):1518–1547.

Voges, W. et al. (1999). The ROSAT all-sky survey bright source catalogue. A&A, 349:389–405.

Voit, G. M. (2005). Tracing cosmic evolution with clusters of galaxies. *Reviews of Modern Physics*, 77(1):207–258.

von der Linden, A., Allen, M. T., Applegate, D. E., Kelly, P. L., Allen, S. W., Ebeling, H., Burchat, P. R., Burke, D. L., Donovan, D., Morris, R. G., Blandford, R., Erben, T., and Mantz, A. (2014). Weighing the Giants - I. Weak-lensing masses for 51 massive galaxy clusters: project overview, data analysis methods and cluster images. MNRAS, 439:2–27.

Walker, A. G. (1935). On the formal comparison of Milne's kinematical system with the systems of general relativity. MNRAS, 95:263–269.

Wallace, D. L. (1958). Asymptotic approximations to distributions. *Ann. Math. Statist.*, 29(3):635–654.

Werner, N., Durret, F., Ohashi, T., Schindler, S., and Wiersma, R. P. C. (2008). Observations of Metals in the Intra-Cluster Medium. Space Sci. Rev., 134(1-4):337–362.

White, M., Cohn, J. D., and Smit, R. (2010). Cluster galaxy dynamics and the effects of large-scale environment. MNRAS, 408(3):1818–1834.

White, S. D. M., Navarro, J. F., Evrard, A. E., and Frenk, C. S. (1993). The baryon content of galaxy clusters: a challenge to cosmological orthodoxy. Nature, 366(6454):429–433.

White, S. D. M. and Rees, M. J. (1978). Core condensation in heavy halos: a two-stage theory for galaxy formation and clustering. MNRAS, 183:341–358.

Will, C. M. (2014). The confrontation between general relativity and experiment. *Living Reviews in Relativity*, 17(1):4.

XRISM Science Team (2020). Science with the X-ray Imaging and Spectroscopy Mission (XRISM). *arXiv e-prints*, page arXiv:2003.04962.

Xu, W., Ramos-Ceja, M. E., Pacaud, F., Reiprich, T. H., and Erben, T. (2018). A new X-ray-selected sample of very extended galaxy groups from the ROSAT All-Sky Survey. A&A, 619:A162.

Yoo, J. and Zaldarriaga, M. (2008). Improved estimation of cluster mass profiles from the cosmic microwave background. Phys. Rev. D, 78(8):083002.

Yoo, J., Zaldarriaga, M., and Hernquist, L. (2010). Lensing reconstruction of cluster-mass cross correlation with cosmic microwave background polarization. Phys. Rev. D, 81(12):123006.

Zaldarriaga, M. and Seljak, U. (1998). Gravitational lensing effect on cosmic microwave background polarization. Phys. Rev. D, 58(2):023003.

Zel'Dovich, Y. B. (1970). Reprint of 1970A&A.....5...84Z. Gravitational instability: an approximate theory for large density perturbations. A&A, 500:13–18.

Zentner, A. R. (2007). The Excursion Set Theory of Halo Mass Functions, Halo Clustering, and Halo Growth. *International Journal of Modern Physics D*, 16(5):763–815.

Zubeldia, Í. and Challinor, A. (2019). Cosmological constraints from Planck galaxy clusters with CMB lensing mass bias calibration. MNRAS, 489:401–419.

Zwicky, F. (1937). On the Masses of Nebulae and of Clusters of Nebulae. ApJ, 86:217.